FELIX KLEIN

ELEMENTARY MATHEMATICS

FROM AN ADVANCED STANDPOINT

ARITHMETIC • ALGEBRA • ANALYSIS

TRANSLATED FROM THE THIRD GERMAN EDITION BY

E. R. HEDRICK
PROFESSOR OF MATHEMATICS
IN THE UNIVERSITY OF CALIFORNIA
AT LOS ANGELES

AND

C. A. NOBLE
PROFESSOR OF MATHEMATICS
IN THE UNIVERSITY OF CALIFORNIA
AT BERKELEY

WITH 125 FIGURES

MACMILLAN AND CO., LIMITED
ST. MARTIN'S STREET, LONDON
1932

51.0

PRINTED IN GERMANY BY THE
SPAMERSCHE BUCHDRUCKEREI • LEIPZIG

Preface to the First Edition.

The new volume which I herewith offer to the mathematical public, and especially to the teachers of mathematics in our secondary schools, is to be looked upon as a first continuation of the lectures *Über den mathematischen Unterricht an den höheren Schulen**, in particular, of those on *Die Organisation des mathematischen Unterrichts*** by *Schimmack* and me, which were published last year by Teubner. At that time our concern was with the different ways in which the problem of instruction can be presented to the mathematician. At present my concern is with developments in the subject matter of instruction. I shall endeavor to put before the teacher, as well as the maturing student, from the view-point of modern science, but in a manner as simple, stimulating, and convincing as possible, both the content and the foundations of the topics of instruction, with due regard for the current methods of teaching. I shall not follow a systematically ordered presentation, as do, for example, Weber and Wellstein, but I shall allow myself free excursions as the changing stimulus of surroundings may lead me to do in the course of the actual lectures.

The program thus indicated, which for the present is to be carried out only for the fields of *Arithmetic, Algebra,* and *Analysis*, was indicated in the preface to Klein-Schimmack (April 1907). I had hoped then that Mr. Schimmack, in spite of many obstacles, would still find the time to put my lectures into form suitable for printing. But I myself, in a way, prevented his doing this by continuously claiming his time for work in another direction upon pedagogical questions that interested us both. It soon became clear that the original plan could not be carried out, particularly if the work was to be finished in a short time, which seemed desirable if it was to have any real influence upon those problems of instruction which are just now in the foreground. As in previous years, then, I had recourse to the more convenient method of *lithographing* my lectures, especially since my present assistant, Dr. Ernst Hellinger, showed himself especially well qualified for this work. One should not underestimate the service which Dr. Hellinger rendered. For it is a far cry from the spoken word of the teacher, influenced as it is by accidental conditions, to the subsequently polished and readable record.

* On the teaching of mathematics in the secondary schools.

** The organization of mathematical instruction.

In precision of statement and in uniformity of explanations, the lecturer stops short of what we are accustomed to consider necessary for a printed publication.

I hesitate to commit myself to still further publications on the teaching of mathematics, at least for the field of *geometry*. I prefer to close with the wish that the present lithographed volume may prove useful by inducing many of the teachers of our higher schools to renewed use of independent thought in determining the best way of presenting the material of instruction. This book is designed solely as such a mental spur, not as a detailed handbook. The preparation of the latter I leave to those actively engaged in the schools. It is an error to assume, as some appear to have done, that my activity has ever had any other purpose. In particular, the *Lehrplan der Unterrichtskommission der Gesellschaft Deutscher Naturforscher und Ärzte** (the so-called "Meraner" *Lehrplan*) is not mine, but was prepared, merely with my cooperation, by distinguished representatives of school mathematics.

Finally, with regard to the method of presentation in what follows, it will suffice if I say that I have endeavored here, as always, to combine geometric intuition with the precision of arithmetic formulas, and that it has given me especial pleasure to follow the historical development of the various theories in order to understand the striking differences in methods of presentation which parallel each other in the instruction of today.

Göttingen, June, 1908

Klein.

Preface to the Third Edition.

After the firm of Julius Springer had completed so creditably the publication of my collected scientific works, it offerred, at the suggestion of Professor Courant, to bring out in book form those of my lecture courses which, from 1890 on, had appeared in lithographed form and which were out of print except for a small reserve stock.

These volumes, whose distribution had been taken over by Teubner, during the last decades were, in the main, the manuscript notes of my various assistants. It was clear to me, at the outset, that I could not undertake a new revision of them without again seeking the help of younger men. In fact I long ago expressed the belief that, beyond a certain age, one ought not to publish independently. One is still qualified, perhaps, to direct in general the preparation of an edition, but is not able to put the details into the proper order and to take into proper account recent advances in the literature. Consequently I accepted the

* Curriculum prepared by the commission on instruction of the Society of German Natural Scientists and Physicians.

offer of Springer only after I was assured that liberal help in this respect would be provided.

These lithographed volumes of lectures fall into two series. The older ones are of special lectures which I gave from time to time, and were prepared solely in order that the students of the following semester might have at hand the material which I had already treated and upon which I proposed to base further work. These are the volumes on *Non-Euclidean Geometry, Higher Geometry, Hypergeometric Functions, Linear Differential Equations, Riemann Surfaces*, and *Number Theory*. In contrast to these, I have published several lithographed volumes of lectures which were intended, from the first, for a larger circle of readers. These are:

a) The volume on *Applications of Differential and Integral Calculus to Geometry,* which was worked up from his manuscript notes by C. H. Müller. This was designed to bridge the gap between the needs of applied mathematics and the more recent investigations of pure mathematicians.

b) and c) Two volumes on *Elementary Mathematics from an Advanced Standpoint,* prepared from his manuscript notes by E. Hellinger. These two were to bring to the attention of secondary school teachers of mathematics and science the significance for their professional work of their academic studies, especially their studies in pure mathematics.

A thoroughgoing revision of the volumes of the second series seemed unnecessary. A smoothing out, in places, together with the addition of supplementary notes, was thought sufficient. With their publication therefore, the initial step is taken. Volumes b), c), a) (in this order) will appear as Parts I, II, III of a single publication bearing the title *Elementary Mathematics from an Advanced Standpoint.* The combining, in this way, of volume a) with volumes b) and c) will meet the approval of all who appreciate the growing significances of applied mathematics for modern school instruction.

Meantime the revision of the volumes of the first series has begun, starting with the volume on *Non-Euclidean Geometry.* But a more drastic recasting of the material will be necessary here if the book is to be a well-rounded presentation, and is to take account of the recent advances of science. So much as to the general plan. Now a few words as to the first part of the *Elementary Mathematics*.

I have reprinted the preface to the 1908 edition of b) because it shows most clearly how the volume came into existence[1]. The second edition (1911), also lithographed, contained no essential changes, and the minor notes which were appended to it are now incorporated into

[1] My co-worker, R. Schimmack, who is mentioned there, died in 1912 at the age of thirty-one years, from a heart attack with which he was seized suddenly, as he sat at his desk.

the text without special mention. The present edition retains[1], in the main, the text of the first edition, including such peculiarities as were incident to the time of its origin. Otherwise it would have been necessary to change the entire articulation, with a loss of homogeneity. But during the sixteen years which have elapsed since the first publication, science has advanced, and great changes have taken place in our school system, changes which are still in progress. This fact is provided for in the appendices which have been prepared, in collaboration with me, by Dr. Seyfarth (Studienrat at the local Oberrealschule). Dr. Seyfarth also made the necessary stylistic changes in the text, and has looked after the printing, including the illustrations, so that I feel sincerely grateful to him. My former co-workers, Messrs. Hellinger and Vermeil, as well as Mr. A. Walther of Göttingen, have made many useful suggestions during the proof reading. In particular, I am indebted to Messrs. Vermeil and Billig for preparing the list of names and the index. The publisher, Julius Springer has again given notable evidence of his readiness to print mathematical works in the face of great difficulties.

Göttingen, Easter, 1924

Klein.

Preface to the English Edition.

Professor *Felix Klein* was a distinguished investigator. But he was also an inspiring teacher. With the rareness of genius, he combined familiarity with all the fields of mathematics and the ability to perceive the mutual relations of these fields; and he made it his notable function, as a teacher, to acquaint his students with mathematics, not as isolated disciplines, but as an integrated living organism. He was profoundly interested in the teaching of mathematics in the secondary schools, both as to the material which should be taught, and as to the most fruitful way in which it should be presented. It was his custom, during many years, at the University of Göttingen, to give courses of lectures, prepared in the interest of teachers and prospective teachers of mathematics in German secondary schools. He endeavored to reduce the gap between the school and the university, to rouse the schools from the lethargy of tradition, to guide the school teaching into directions that would stimulate healthy growth; and also to influence university attitude and teaching toward a recognition of the normal function of the secondary school, to the end that mathematical education should be a continuous growth.

These lectures of Professor *Klein* took final form in three printed volumes, entitled *Elementary Mathematics from an Advanced Standpoint.*

[1] New comments are placed in brackets.

They constitute an invaluable work, serviceable alike to the university teacher and to the teacher in the secondary school. There is, at present, nothing else comparable with them, either with respect to their skilfully integrated material, or to the fascinating way in which this material is discussed. This English volume is a translation of Part I of the above work. Its preparation is the result of a suggestion made by Professor *Courant*, of the University of Göttingen. It is the expression of a desire to serve the need, in English speaking countries, of actual and prospective teachers of mathematics; and it appears with the earnest hope that, in a rather free translation, something of the spirit of the original has been retained.

The Translators.

Contents

Introduction . . . .

First Part: Arithme

I. Calculating with Natural Numbers . . . . .
1. Introduction of Numbers in the Schools. .
2. The Fundamental Laws of Reckoning . .
3. The Logical Foundations of Operations with
4. Practice in Calculating with Integers . . .

II. The First Extension of the Notion of Numb
1. Negative Numbers . . . .
2. Fractions . . . .
3. Irrational Numbers . . . .

III. Concerning Special Properties of Integers .

IV. Complex Numbers . . . .
1. Ordinary Complex Numbers . . . .
2. Higher Complex Numbers, especially Quate
3. Quaternion Multiplication—Rotation and E
4. Complex Numbers in School Instruction. .
Concerning the Modern Development and t
Mathematics . . . .

Second Part: Alge

I. Real Equations with Real Unknowns . . .
1. Equations with one parameter . . . . . .
2. Equations with two parameters . . . . .
3. Equations with three parameters λ, μ, ν. .

II. Equations in the field of complex quantitie
A. The fundamental theorem of algebra . . .
B. Equations with a complex parameter. . .
1. The "pure" equation . . . .
2. The dihedral equation . . . .
3. The tetrahedral, the octahedral, and th
4. Continuation: Setting up the Normal E
5. Concerning the Solution of the Normal
6. Uniformization of the Normal Irrational
cendental Functions . . . .
7. Solution in Terms of Radicals . . . .
8. Reduction of Genral Equations to Norn

Third Part: Analysis

Page
I. Logarithmic and Exponential Functions 144
1. Systematic Account of Algebraic Analysis 144
2. The Historical Development of the Theory 146
3. The Theory of Logarithms in the Schools 155
4. The Standpoint of Function Theory 156

II. The Goniometric Functions 162
1. Theory of the Goniometric Functions 162
2. Trigonometric Tables 169
A. Purely Trigonometric Tables 170
B. Logarithmic—Trigonometric Tables 172
3. Applications of Goniometric Functions 175
A. Trigonometry, in particular, spherical trigonometry 175
B. Theory of small oscillations, especially those of the pendulum . 186
C. Representation of periodic functions by means of series of goniometric functions (trigonometric series) 190

III. Concerning Infinitesimal Calculus Proper 207
1. General Considerations in Infinitesimal Calculus 207
2. TAYLORS Theorem 223
3. Historical and Pedagogical Considerations 234

Supplement

I. Transcendence of the Numbers e and π 237

II. The Theory of Assemblages 250
1. The Power of an Assemblage 251
2. Arrangement of the Elements of an Assemblage 262

Index of Names 269

Index of Contents 271

Introduction

In recent years[1], a far reaching interest has arisen among university teachers of mathematics and natural science directed toward a suitable training of candidates for the higher teaching positions. This is really quite a new phenomenon. For a long time prior to its appearance, university men were concerned exclusively with their sciences, without giving a thought to the needs of the schools, without even caring to establish a connection with school mathematics. What was the result of this practice? The young university student found himself, at the outset, confronted with problems which did not suggest, in any particular, the things with which he had been concerned at school. Naturally he forgot these things quickly and thoroughly. When, after finishing his course of study, he became a teacher, he suddenly found himself expected to teach the traditional elementary mathematics in the old pedantic way; and, since he was scarcely able, unaided, to discern any connection between this task and his university mathematics, he soon fell in with the time honored way of teaching, and his university studies remained only a more or less pleasant memory which had no influence upon his teaching.

There is now a movement to abolish this double discontinuity, helpful neither to the school nor to the university. On the one hand, there is an effort to impregnate the material which the schools teach with new ideas derived from modern developments of science and in accord with modern culture. We shall often have occasion to go into this. On the other hand, the attempt is made to take into account, in university instruction, the needs of the school teacher. And it is precisely in such comprehensive lectures as I am about to deliver to you that I see one of the most important ways of helping. I shall by no means address myself to beginners, but I shall take for granted that you are all acquainted with the main features of the chief fields of mathematics. I shall often talk of problems of algebra, of number theory, of function theory, etc., without being able to go into details. You must, therefore, be moderately familiar with these fields, in order to follow me. My task will always be to show you the *mutual connection between problems in*

[[1] Attention is again drawn to the fact that the wording of the text is, almost throughout, that of the lithographed volume of 1908 and that comments which refer to later years have been put into the appendices.]

the various fields, a thing which is not brought out sufficiently in the usual lecture course, and more especially to emphasize the relation of these problems to those of school mathematics. In this way I hope to make it easier for you to acquire that ability which I look upon as the real goal of your academic study: the ability to draw (in ample measure) from the great body of knowledge there put before you a living stimulus for your teaching.

Let me now put before you some documents of recent date which give evidence of widespread interest in the training of teachers and which contain valuable material for us. Above all I think here of the addresses given at the last *Meeting of Naturalists* held September 16, 1907, in Dresden, to which body we submitted the Proposals for the Scientific Training of Prospective Teachers of Mathematics and Science of the Committee on Instruction of the Society of German Naturalists and Physicians. You will find these Proposals as the last section in the Complete Report of this Committee[1] which, since 1904, has been considering the entire complex of questions concerning instruction in mathematics and natural science and has now ended its activity; I urge you to take notice, not only of these Proposals, but also of the other parts of this very interesting report. Shortly after the Dresden meeting there occurred a similar debate at the Meeting of German Philologists and Schoolmen in Basel, September 25, in which, to be sure, the mathematical-scientific reform movement was discussed only as a link in the chain of parallel movements occurring in philological circles. After a report by me concerning our aims in mathematical-natural science reform there were addresses by P. Wendland (Breslau) on questions in *Archeology*, Al. Brandl (Berlin) on *modern languages* and, finally, *Ad. Harnack* (Berlin) on *History and religion*. These four addresses appeared together in one broschure[2] to which I particulary refer you. I hope that this auspicious beginning will develop into further cooperation between our scientists and the philologists, since it will bring about friendly feeling and mutual understanding between two groups whose relations have been unsympathetic even if not hostile. Let us endeavor always to foster such good relations even if we do among ourselves occasionally drop a critical word about the philologists, just as they may about us. Bear in mind that you will later be called upon in the schools to work together with the philologists for the common good and that this requires mutual understanding and appreciation.

Along with this evidence of efforts which reach beyond the borders of our field, I should like to mention a few books which aim in the

[1] *Die Tätigkeit der Unterrichtskommission der Gesellschaft deutscher Naturforscher und Ärzte*, edited by A. Gutzmer. Leipzig and Berlin, 1908.

[2] *Universität und Schule.* Addresses delivered by F. Klein, P. Wendland, Al. Brandl, Ad. Harnack. Leipzig 1907.

same direction in the mathematical field and which will therefore be important for these lectures. Three years ago I gave, for the first time, a course of lectures with a similar purpose. My assistant at that time, R. Schimmack, worked the material up and the first part has recently appeared in print[1]. In it are considered the different kinds of schools, including the university, the conduct of mathematical instruction in them, the interests that link them together, and other similar matters. In what follows I shall from time to time refer to things which appear there without repeating them. This makes it possible for me to extend somewhat those considerations. That volume concerns itself with the organization of school instruction. I shall now consider the mathematical content of the material which enters into that instruction. If I frequently advert to the actual conduct of instruction in the schools, my remarks will be based not merely upon indefinite pictures of how the thing might be done or even upon dim recollections of my own school days; for I am constantly in touch with Schimmack, who is now teaching in the Göttingen gymnasium and who keeps me informed as to the present state of instruction, which has, in fact, advanced substantially beyond what it was in earlier years. During this winter semester I shall discuss "the three great A's", that is arithmetic, algebra, and analysis, withholding geometry for a continuation of the course during the coming summer. Let me remind you that, in the language of the secondary schools, these three subjects are classed together as arithmetic, and that we shall often note deviations in the terminology of the schools as compared with that at the universities. You see, from this small illustration, that only living contact can bring about understanding.

As a second reference I shall mention the three volume *Enzyklopädie der Elementarmathematik* by H. Weber and J. Wellstein, the work which, among recent publications, most nearly accords with my own tendencies. For this semester, the first volume, *Enzyklopädie der elementaren Algebra und Analysis*, prepared by H. Weber[2], will be the most important. I shall indicate at once certain striking differences between this work and the plan of my lectures. In Weber-Wellstein, the entire structure of elementary mathematics is built up systematically and logically in the mature language of the advanced student. No account is taken of how these things actually may come up in school instruction. The presentation in the schools, however, should be psychological and not systematic. The teacher so to speak, must be a diplomat. He must take account of the psychic processes in the boy in order to grip his interest;

[1] Klein, F., *Vorträge über den mathematischen Unterricht an höheren Schulen.* Prepared by von R. Schimmack. Part 1. *Von der Organisation des mathematischen Unterrichts.* Leipzig 1907. This book is referred to later as "Klein-Schimmack".

[2] Second edition. Leipzig 1906. [Fourth edition, 1922, revised by P. Epstein. — Referred to as "Weber-Wellstein I".

and he will succeed only if he presents things in a form intuitively comprehensible. A more abstract presentation will be possible only in the upper classes. For example: The child cannot possibly understand if numbers are explained axiomatically as abstract things devoid of content, with which one can operate according to formal rules. On the contrary, he associates numbers with concrete images. They are numbers of nuts, apples, and other good things, and in the beginning they can be and should be put before him only in such tangible form. While this goes without saying, one should—mutatis mutandis—take it to heart, that in all instruction, even in the university, mathematics should be associated with everything that is seriously interesting to the pupil at that particular stage of his development and that can in any way be brought into relation with mathematics. It is just this which is back of the recent efforts to give prominence to applied mathematics at the university. This need has never been overlooked in the schools so much as it has at the university. It is just this psychological value which I shall try to emphasize especially in my lectures.

Another difference between Weber-Wellstein and myself has to do with defining the content of school mathematics. Weber and Wellstein are disposed to be conservative, while I am progressive. These things are thoroughly discussed in Klein-Schimmack. We, who are called the reformers, would put the function concept at the very center of instruction, because, of all the concepts of the mathematics of the past two centuries, this one plays the leading role wherever mathematical thought is used. We would introduce it into instruction as early as possible with constant use of the graphical method, the representation of functional relations in the xy system, which is used today as a matter of course in every practical application of mathematics. In order to make this innovation possible, we would abolish much of the traditional material of instruction, material which may in itself be interesting, but which is less essential from the standpoint of its significance in connection with modern culture. Strong development of space perception, above all, will always be a prime consideration. In its upper reaches, however, instruction should press far enough into the elements of infinitesimal calculus for the natural scientist or insurance specialist to get at school the tools which will be indispensable to him. As opposed to these comparatively recent ideas, Weber-Wellstein adheres essentially to the traditional limitations as to material. In these lectures I shall of course be a protagonist of the new conception.

My third reference will be to a very stimulating book: *Didaktik und Methodik des Rechnens und der Mathematik*[1] by Max Simon, who like

[1] Second edition, Munich 1908. Separate reprint from Baumeister's *Handbuch der Erziehungs- und Unterrichtslehre für höhere Schulen*, first edition, 1895.

Weber and Wellstein is at Strassburg. Simon is often in agrement with our views, but he sometimes takes the opposite standpoint; and inasmuch as he is a very subjective, temperamental, personality he often clothes these contrasting views in vivid words. To give one example, the proposals of the Committee on Instruction of the Natural Scientists require an hour of geometric propaedeutics in the second year of the gymnasium, whereas at the present time this usually begins in the third year. It has long been a matter of discussion which plan is the better; and the custom in the schools has often changed. But Simon declares the position taken by the Commission, which, mind you, is at worst open to argument, to be "worse than a crime", and that without in the least substantiating his judgment. One could find many passages of this sort. As a precursor of this book I might mention Simon's *Methodik der elementaren Arithmetik in Verbindung mit algebraischer Analysis*[1].

After this brief introduction let us go over to the subject proper, which I shall consider under three headings, as above indicated.

---

[1] Leipzig 1906.

First Part

Arithmetic

I. Calculating with Natural Numbers

We begin with the foundation of all arithmetic, calculation with positive integers. Here, as always in the course of these lectures, we first raise the question as to how these things are handled in the schools; then we shall proceed to the question as to what they imply when viewed from an advanced standpoint.

1. Introduction of Numbers in the Schools

I shall confine myself to brief suggestions. These will enable you to recall how you yourselves learned your numbers. In such an exposition it is, of course, not my purpose to induct you into the practice of teaching, as is done in the Seminars of the secondary schools. I shall merely exhibit the material upon which we shall base our critique.

The problem of teaching children the properties of integers and how to reckon with them, and of leading them on to complete mastery, is very difficult and requires the labor of several years, from the first school year until the child is ten or eleven years old. The manner of instruction as it is carried on in this field in Germany can perhaps best be designated by the words *intuitive* and *genetic*, i. e., the entire structure is gradually erected on the basis of familiar, concrete things, in marked contrast to the customary *logical* and *systematic* method at the university.

One can divide up this material of instruction roughly as follows. The entire first year is occupied with the integers from 1 to 20, the first half being devoted to the range 1 to 10. The integers appear at first as numbered pictures of points or as arrays of all sorts of objects familiar to the children. Addition and multiplication are then presented by intuitional methods, and are fixed in mind.

In the second stage, the integers from 1 to 100 are considered and the Arabic numerals, together with the notion of positional value and the decimal system, are introduced. Let us note, incidentally, that the name "Arabic numerals", like so many others in science, is a misnomer. This form of writing was invented by the Hindus, not by the Arabs. Another principal aim of the second stage is knowledge of the multi-

plication table. One must know what 5×7 or 3×8 is in one's sleep, so to speak. Consequently the pupil must learn the multiplication table by heart to this degree of thoroughness, to be sure only after it has been made clear to him visually with concrete things. To this end the *abacus* is used to advantage. It consists, as you all know, of 10 wires stretched one above another, upon each of which there are strung ten movable beads. By sliding these beads in the proper way, one can read off the result of multiplication and also its decimal form.

The third stage, finally, teaches calculation with numbers of more than one digit, based on the known simple rules whose general validity is evident, or should be evident, to the pupil. To be sure, this evidence does not always enable the pupil to make the rules completely his own; they are often instilled with the authoritative dictum: "It is thus and so, and if you don't know it yet, so much the worse for you!"

I should like to emphasize another point in this instruction which is usually neglected in university teaching. It is that the application of numbers to practical life is strongly emphasized. From the beginning, the pupil is dealing with numbers taken from real situations, with coins, measures, and weights; and the question, "*What does it cost?*", which is so important in daily life, forms the pivot of much of the material of instruction. This plan rises soon to the stage of problems, when deliberate thought is necessary in order to determine what calculation is demanded. It leads to the problems in proportion, alligation, etc. To the words *intuitive* and *genetic*, which we used above to designate the character of this instruction, we can add a third word, *applications*.

We might summarize the purpose of the number work by saying: *It aims at reliability in the use of the rules of operation, based on a parallel development of the intellectual abilities involved, and without special concern for logical relations.*

Incidentally, I should like to direct your attention to a contrast which often plays a mischievous role in the schools, viz., the contrast between the university-trained teachers and those who have attended normal schools for the preparation of elementary school teachers. The former displace the latter, as teachers of arithmetic, during or after the sixth school year, with the result that a regrettable discontinuity often manifests itself. The poor youngsters must suddenly make the acquaintance of new expressions, whereas the old ones are forbidden. A simple example is the different multiplication signs, the $\times$ being preferred by the elementary teacher, the point by the one who has attended the university. Such conflicts can be dispelled, if the more highly trained teacher will give more heed to his colleague and will try to meet him on common ground. That will become easier for you, if you will realize what high regard one must have for the performance of the elementary school teachers. Imagine what methodical training is ne-

cessary to indoctrinate over and over again a hundred thousand stupid, unprepared children with the principles of arithmetic! Try it with your university training; you will not have great success!

Returning, after this digression, to the material of instruction, we note that after the third year of the gymnasium*, and especially in the fourth year, arithmetic begins to take on the more aristocratic dress of mathematics, for which the transition to operations with letters is characteristic. One designates by a, b, c, or x, y, z any numbers, at first only positive integers, and applies the rules and operations of arithmetic to the numbers thus symbolized by letters, whereby the numbers are devoid of concrete intuitive content. This represents such a long step in abstraction that one may well declare that real mathematics begins with operations with letters. Naturally this transition must not be accomplished rapidly. The pupils must accustom themselves gradually to such marked abstraction.

It seems unquestionably necessary that, for this instruction, the teacher should know thoroughly the logical laws and foundations of reckoning and of the theory of integers.

2. The Fundamental Laws of Reckoning

Addition and multiplication were familiar operations long before any one inquired as to the fundamental laws governing these operations. It was in the twenties and thirties of the last century that particularly English and French mathematicians formulated the fundamental properties of the operations, but I will not enter into historical details here. If you wish to study these, I recommend to you, as I shall often do, the great *Enzyklopädie der Mathematischen Wissenschaften mit Einschluß ihrer Anwendungen*[1], and also the French translation: *Encyclopédie des Sciences mathématiques pures et appliquées*[2] which bears in part the character of a revised and enlarged edition. If a school library has only one mathematical work, it ought to be this encyclopedia, for through it the teacher of mathematics would be placed in position to continue his work in any direction that might interest him. For us, at this place, the article of interest is the first one in the first volume[3] H. Schubert: *"Grundlagen der Arithmetik"*, of which the translation into French is by Jules Tannery and Jules Molk.

* The German gymnasium is a nine-year secondary school, following a four-year preparatory school. Hence the third year of the gymnasium is the student's seventh school year.

[1] Leipzig (B. G. Teubner) from 1908 on. Volume I has appeared complete, Volumes II–VI are nearing completion.

[2] Paris (Gauthur-Villars) and Leipzig (Teubner) from 1904 on; unfortunately the undertaking had to be abandoned after the death of its editor J. Molk (1914).

[3] *Arithmetik und Algebra*, edited by W. Fr. Meyer (1896–1904); in the French edition, the editor was J. Molk.

Going back to our theme, I shall enumerate the *five fundamental laws* upon which *addition* depends:

1. *$a + b$ is always again a number*, i. e., *addition is always possible* (in contrast to subtraction, which is not always possible in the domain of positive integers).
2. *$a + b$ is one-valued.*
3. *The associative law holds:*

$$(a + b) + c = a + (b + c),$$

so that one may omit the parentheses entirely.

4. *The commutative law holds:*

$$a + b = b + a.$$

5. *The monotonic law holds:*

$$\text{If } b > c, \text{ then } a + b > a + c.$$

These properties are all obvious immediately if one recalls the process of counting; but they must be formally stated in order to justify logically the later developments.

For multiplication there are *five exactly analogous laws:*

1. *$a \cdot b$ is always a number.*
2. *$a \cdot b$ is one-valued.*
3. *Associative law:* $a \cdot (b \cdot c) = (a \cdot b) \cdot c = a \cdot b \cdot c$.
4. *Commutative law:* $a \cdot b = b \cdot a$.
5. *Monotonic law: If $b > c$, then $a \cdot b > a \cdot c$.*

Multiplication together with addition obeys also the following law.

6. *Distributive law:*

$$a \cdot (b + c) = a \cdot b + a \cdot c.$$

It is easy to show that all elementary reckoning can be based upon these eleven laws. It will be sufficient to illustrate this fact by a simple example, say the multiplication of 7 and 12. From the distributive law we have:

$$7 \cdot 12 = 7 \cdot (10 + 2) = 70 + 14,$$

and if we separate 14 into $10 + 4$ (carrying the tens), we have, by the associative law of addition,

$$70 + (10 + 4) = (70 + 10) + 4 = 80 + 4 = 84.$$

You will recognize in this procedure the steps of the usual decimal reckoning. It would be well for you to construct for yourselves more complicated examples. We might summarize by saying *that ordinary reckoning with integers consists in repeated use of the eleven fundamental laws together with the memorized results of the addition and multiplication tables.*

But where does one use the monotonic laws? In ordinary formal reckoning, to be sure, they are superfluous, but not in certain other

problems. Let me remind you of the process called *abridged multiplication and division* with decimal numbers[1]. That is a thing of great practical importance which unfortunately is too little known in the schools, as well as among university students, although it is sometimes mentioned in the second year of the gymnasium. As an example, suppose that one wished to compute $567 \cdot 134$, and that the units digit in each number was of questionable accuracy, say as a result of physical measurement. It would be unnecessary work, then, to determine the product *exactly*, since one could not guarantee an exact result. It is, however, important to know the *order* of *magnitude* of the product, i. e., to know between which tens or between which hundreds the exact value lies. The monotonic law supplies this estimate at once; for it follows by that law that the desired value lies between $560 \cdot 134$ and $570 \cdot 134$ or between $560 \cdot 130$ and $570 \cdot 140$. I leave to you the carrying out of the details; at least you see that *the monotonic law is continually used in abridged reckoning.*

A systematic exposition of these fundamental laws is, of course, not to be thought of in the secondary schools. After the pupils have gained a concrete understanding and a secure mastery of reckoning with numbers, and are ready for the transition to operations with letters, the teacher should take the opportunity to state, at least, the associative, commutative, and distributive laws and to illustrate them by means of numerous obvious numerical examples.

3. The Logical Foundations of Operations with Integers

While instruction in the schools will naturally not rise to still more difficult questions, present mathematical investigation really begins with the question: *How does one justify the above-mentioned fundamental laws, how does one account for the notion of number at all?* I shall try to explain this matter in accordance with the announced purpose of these lectures to endeavor to get new light upon school topics by looking at them from another point of view. I am all the more willing to do this because these modern thoughts crowd in upon you from all sides during your academic years, but not always accompanied by any indication of their psychological significance.

First of all, so far as the notion of number is concerned, it is very difficult to discover its origin. Perhaps one is happiest if one decides to ignore these most difficult things. For more complete information as to these questions, which are so earnestly discussed by the philosophers, I must refer you to the article, already mentioned, in the French encyclopedia, and I shall confine myself to a few remarks. A widely accepted belief is that the notion of number is closely connected with the notion of time, with temporal succession. The philosopher Kant

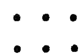
[1] The monotonic laws will be used later, also, in the theory of irrational numbers.

and the mathematician Hamilton represent this view. Others think that number has more to do with space perception. They base the notion of number upon the *simultaneous perception of different objects which are near each other.* Still others see, in number concepts, the expression of a *peculiar faculty of the mind* which exists independently of, and coordinate with, or even above, perception of space and time. I think that this conception would be well characterized by quoting from Faust the lines which Minkowski, in the preface of his book on *Diophantine Approximation,* applies to numbers:

"Göttinnen thronen hehr in Einsamkeit,
Um sie kein Ort, noch weniger eine Zeit."

While this problem involves primarily questions of psychology and epistemology, the justification of our eleven laws, at least the recent researches regarding their compatibility, implies questions of logic. We shall distinguish the following four points of view.

1. According to the first of these, best represented perhaps by Kant, the rules of reckoning are immediate necessary results of perception, whereby this word is to be understood, in its broadest sense, as "inner perception" or intuition. It is not to be understood by this that mathematics rests throughout upon experimentally controllable facts of external experience. To mention a simple example, the commutative law is established by examining the accompanying picture, which consists of two rows of three points each, that is, $2 \cdot 3 = 3 \cdot 2$. If the objection is raised that in the case of only moderately large numbers, this immediate perception would not suffice, the reply is that we call to our assistance the *theorem of mathematical induction. If a theorem holds for small numbers, and if an assumption of its validity for a number n always insures its validity for $n + 1$, then it holds generally for every number.* This theorem, which I consider to be really an intuitive truth, carries us over the boundary where sense perception fails. This standpoint is more or less that of Poincaré in his well known philosophical writings.



If we would realize the significance of this question as to the source of the validity of our eleven fundamental rules of reckoning, we should remember that, along with arithmetic, mathematics as a whole rests ultimately upon them. Thus it is not asserting too much to say, that, according to the conception of the rules of reckoning which we have just outlined, *the security of the entire structure of mathematics rests upon intuition, where this word is to be understood in its most general sense.*

2. The second point of view is a modification of the first. According to it, one tries to separate the eleven fundamental laws into a larger number of shorter steps of which one need take only the simplest directly from intuition, while the remainder are deduced from these by rules of logic without any further use of intuition. Whereas, before,

the possibility of logical operation began *after* the eleven fundamental laws had been set up, it can start earlier here, after the simpler ones have been selected. *The boundary between intuition and logic is displaced in favor of the latter.* Hermann Grassmann did pioneer work in this direction in his *Lehrbuch der Arithmetik*[1] in 1861. As an example from it, I mention merely that the commutative law can be derived from the associative law by the aid of the principle of mathematical induction. Because of the precision of his presentation, one might place by the side of this book of Grassmann one by the Italian Peano, *Arithmetices principia nova methodo exposita*[2]. Do not assume, however, because of this title, that the book was written in Latin! It is written in a peculiar symbolic language designed by the author to display each logical step of the proof and emphasize it as such. Peano wishes to have a guarantee in this way, that he is making use only of the principle which he specifically mentions, with nothing whatever coming from intuition. He wishes to avoid the danger that countless uncontrollable associations of ideas and reminders of perception might creep in if he used our ordinary language. Note, too, that Peano is the leader of an extensive Italian school which is trying in a similar way to separate into small groups the premises of each individual branch of mathematics, and, with the aid of such a symbolic language, to investigate their exact logical connections.

3. We come now to a *modern extension of these ideas*, which has, moreover, been influenced by Peano. I refer to that treatment of the foundations of arithmetic which puts the theory of point sets into the foreground. You will be able to form a notion of the wide range of the idea of a point set if I tell you that the totality of all integers, as well as that of all points on a line segment, are special examples of point sets. Georg Cantor, as is generally known, was the first to make this general idea the object of orderly mathematical speculation. The *theory of point sets*, which he created, is now claiming the profound attention of the younger generation of mathematicians. Later I shall endeavor to give you a cursory view of this subject. For the present, it is sufficient to characterize as follows the tendency of the new foundation of arithmetic which have been based upon it: *The properties of integers and of operations with them are to be deduced from the general properties and abstract relations of point sets*, in order that the foundation may be as sound and general as possible.

---

[1] With the addition to the title *"für höhere Lehranstalten"* (Berlin 1861). The corresponding chapters are reprinted in H. Grassmann's *Gesammelten mathematischen und physikalischen Werken* (edited by F. Engel), Vol. II, 1, pp. 295–349. Leipzig 1904.

[2] Augustae Taurinorum. Torino 1889. [There is a more comprehensive presentation in Peano's *Formulaire de Mathématiques* (1892–1899).

One of the pioneers along this path was Richard Dedekind, who, in his small but important book *Was sind und was sollen die Zahlen?*[1], attempted such a foundation for integers. *H. Weber* inclines to this point of view in the first part of Weber-Wellstein, volume I (See p. 3). To be sure, the deduction is quite abstract and offers, still, certain grave difficulties, so that Weber, in an Appendix to Volume III[2], gave a more elementary presentation, using only finite point sets. In later editions, this appendix is incorporated into Volume I. Those of you who are interested in such questions are especially referred to this presentation.

4. Finally, I shall mention the *purely formal theory of numbers*, which, indeed, goes back to Leibniz and which has recently been brought into the foreground again by Hilbert. His address *Über die Grundlagen der Logik und Arithmetik** at the Heidelberg Congress in 1904 is important for arithemtic[3]. His fundamental conception is as follows: Once one has the eleven fundamental rules of reckoning, one can operate with the letters $a, b, c, \ldots$, which actually represent arbitrary integers, without bearing in mind that they have a real meaning as numbers. In other words: let a, $b, c, \ldots$, be things devoid of meaning, or things of whose meaning we know nothing; let us agree only that one may combine them according to those eleven rules, but that these combinations need not have any real known meaning. Obviously one can than operate with $a, b, c, \ldots$, precisely as one ordinarily does with actual numbers. Only the question arises here *whether these operations could lead one to contradictions*. Now ordinarily one says that intuition shows us the existence of numbers for which these eleven laws hold, and that it is consequently impossible for contradictions to lurk in these laws. But in the present case, where we are not thinking of the symbols as having definite meaning, such an appeal to perception is not permissible. *In fact, there arises the entirely new problem, to prove logically that no operations with our symbols which are based on the eleven fundamental laws can ever lead to a contradiction, i. e., that these eleven laws are consistent, or compatible.* While we were discussing the first point of view, we took the position that the certainty of mathematics rests upon the existence of intuitional things which fit its theorems. The adherents of this formal standpoint, on the other hand, must hold that *the certainty of mathematics rests upon the possibility of showing that the fundamental laws considered formally and without reference to their intuitional content, constitute a logically consistent system.*

---

[1] Braunschweig 1888; third edition 1911.

[2] *Angewandte Elementarmathematik.* Revised by H. Weber, J. Wellstein, R. H. Weber. Leipzig 1907.

* *On the foundations of logic and arithmetic.*

[3] *Verhandlungen des 3. internationalen Mathematikerkongresses in Heidelberg* August 8—13, 1904, p. 174 et seq., Leipzig 1905.

I shall close this discussion with the following remarks:

a) Hilbert indicated all of these points of view in his Heidelberg address, but he followed none of them through completely. Afterwards he pushed them somewhat farther in a course of lectures, but then abandoned them. We can thus say that *here is a field for investigation*[1].

b) The tendency to crowd intuition completely off the field and to attain to really *pure* logical investigations seems to me not completely feasible. It seems to me that *one must retain something, albeit a minimum, of intuition.* One must always use a certain intuition in the most abstract formulation with the symbols one uses in operations, in order to recognize the symbols again, even if one thinks only about the shape of the letters.

c) Let us even assume that the proposed problem has been solved in a way free from objection, that the compatibility of the eleven fundamental laws has been proved logically. Precisely at this point an opening is offered for a remark which I should like to make with the utmost emphasis. One must *see clearly that the real arithmetic, the theory of actual integers, is neither established, nor can ever be established, by considerations of this nature.* It is impossible to show in a purely logical way that the laws whose consistency is established in that manner are actually valid for the numbers with which we are intuitionally familiar; that the undefined things of which we speak, and the operations which we apply to them, can be identified with actual numbers and with the processes of addition and multiplication in their intuitively clear significance. What is accomplished is, rather, that the tremendous *problem of building the foundations of arithmetic, unassailable in its complexity, is split into two parts,* and that the first, the purely logical problem, *the setting up of independent fundamental laws or axioms and the investigation of them as to independence and consistency* has been made available to study. The second, the more epistemological part of the problem, which has to do with the justification for the *application of these laws to actual conditions,* is not even touched, although it must of course be solved also if one will really build the foundations of arithmetic. This second part presents, in itself, an extremely profound problem, whose difficulties lie in the general field of epistemology. I can characterize its standing most clearly perhaps, by the somewhat paradoxical remark that anyone who tolerates only pure logic in investigations in pure mathematics must, to be consistent, look upon the second part of the problem of the foundation of arithmetic, and hence upon arithmetic itself, as belonging to *applied* mathematics.

[[1] Concerning more recent developments in these investigations, see the preceding footnote.]

I have felt obliged to go into detail here very carefully, in as much as misunderstandings occur so often at this point, because people simply overlook the existence of the second problem. This is by no means the case with Hilbert himself, and neither my disagreement nor my agreement with him is a warranted conclusion if it be based on such an assumption.

Thomae of Jena, coined the neat expression "thoughtless thinkers" for those persons who confine themselves exclusively to these abstract investigations concerning things that are devoid of meaning, and to theorems that tell nothing, and who forget not only that second problem but often also all the rest of mathematics. This facetious term cannot apply, of course, to people who carry on those investigations alongside of many others of a different sort.

In connection with this brief survey of the foundation of arithmetic, I shall bring to your notice a few general matters. Many have thought that one could, or that one indeed must, teach all mathematics deductively *throughout*, by starting with a definite number of axioms and deducing everything from these by means of logic. This method, which some seek to maintain upon the authority of Euclid, certainly does not correspond to the historical development of mathematics. In fact, mathematics has grown like a tree, which does not start at its tiniest rootlets and grow merely upward, but rather sends its roots deeper and deeper at the same time and rate that its branches and leaves are spreading upward. Just so – if we may drop the figure of speech –, mathematics began its development from a certain standpoint corresponding to normal human understanding, and has progressed, from that point, according to the demands of science itself and of the then prevailing interests, now in the one direction toward new knowledge, now in the other through the study of fundamental principles. For example, our standpoint today with regard to foundations is different from that of the investigators of a few decades ago; and what we today would state as ultimate principles, will certainly be outstripped after a time, in that the latest truths will be still more meticulously analyzed and referred back to something still more general. *We see, then, that as regards the fundamental investigations in mathematics, there is no final ending, and therefore, on the other hand, no first beginning, which could offer an absolute basis for instruction.*

Still another remark concerning the relation between the logical and the intuitional handling of mathematics, between pure and applied mathematics. I have already emphasized the fact that, in the schools, applications accompany arithmetic from the beginning, that the pupil learns not only to understand the rules, but to do something with them. And it should always be so in the teaching of mathematics! Of course, the logical connections, one might say *the rigid skeleton in the mathematical*

organism, must remain, in order to give it its peculiar trustworthiness. But the living thing in mathematics, its most important stimulus, its effectiveness in all directions, depends entirely upon the applications, i. e., upon the mutual relations between those purely logical things and all other domains. To banish applications from mathematics would be comparable to seeking the essence of the living animal in the skeleton alone, without considering muscles, nerves and tissues, instincts, in short, the very life of the animal.

In scientific *investigation* there is often, to be sure, a *division of labor* between pure and applied science, but when this happens, provision must be made otherwise for maintaining their connection if conditions are to remain sound. In any case, and this should be especially emphasized here, *for the school such a division of labor, such a fareaching specialization of the individual teacher, is not possible.* To put the matter crassly, imagine that at a certain school a teacher is appointed who treats numbers only as meaningless symbols, a second teacher who knows how to bridge the gap from these empty symbols to actual numbers, a third, a fourth, a fifth, finally, who understands the application of these numbers to geometry, to mechanics, and to physics; and that these different teachers are all turned lose upon the pupils. You see that such an organization of teaching is impossible. In this way, the things could not be brought to the comprehension of the pupils, neither would the individual teachers be able even to understand each other. The needs of school instruction itself require precisely a certain many sidedness of the individual teacher, a comprehensive orientation in the field of pure and applied mathematics, in the broadest sense, and include thus a desirable remedy against a too extensive splitting up of science.

In order to give a practical turn to the last remarks I refer again to our above mentioned *Dresden Proposals*. There we recommend outright that applied mathematics, which since 1898 has been a special subject in the examination for prospective teachers, be made a required part in all normal mathematical training, so that competence to teach pure and applied mathematics should always be combined. In addition to this, it should be noted that, in the Meran Curriculum[1] of the Commission of Instruction, the following three tasks are announced as the *purpose of mathematical instruction in the last school year:*

1. *A scientific survey of the systematic structure of mathematics.*

2. *A certain degree of skill in the complete handling, numerical and graphical, of problems.*

---

[1] *Reformvorschläge für den mathematischen und naturwissenschaftlichen Unterricht, überreicht der Versammlung der Naturforscher und Ärzte zu Meran.* Leipzig, 1905. — See also a reprint in the *Gesamtbericht der Kommission*, p. 93, as well as in Klein-Schimmack, p. 208.

3. *An appreciation of the significance of mathematical thought for a knowledge of nature and for modern culture.*

All these formulations I approve with deep conviction.

4. Practice in Calculating with Integers

Turning from discussions which have been chiefly abstract, let us give our attention to more concrete things by considering the *carrying out of numerical calculation.* As suitable literature for collateral reading, I should mention first of all, the article on *Numerisches Rechnen* by R. Mehnicke[1] in the *Enzyclopädie.* I can best give you a general view of the things that belong here by giving a brief account of this article. It is divided into two parts: A. *Die Lehre vom genauen Rechnen**, and B. *Die Lehre vom genäherten Rechnen***. Under A occur all methods for simplifying exact calculation with large integers. *Convenient devices for calculating, tables of products and squares,* and in particular, *calculating machines,* which we shall discuss soon. Under B, on the other hand, one finds a discussion of the methods and devices for all calculating in which only the *order of magnitude of the result* is important, especially *logarithmic tables and allied devices,* the *slide rule,* which is only an expecially well-arranged graphical logarithmic table; finally, also, the numerous important *graphical methods.* In addition to this reference I can recommend the little book by J. Lüroth, *Vorlesungen über numerisches Rechnen*[2]***, which, written in agreeable form by a master of the subject, gives a rapid survey of this field.

From the many topics that have to do with calculating with integers, I shall select for discussion only the calculating machine, which you will find in use, in a great variety of ingenious forms, by the larger banks and business houses, and which is really of the greatest practical significance. We have in our mathematical collection one of the most widely used types, the "Brunsviga", manufactured by the firm Brunsviga-Maschinenwerke Grimme, Natalis & Co. A.-G. in Braunschweig. The design originated with the Swedish engineer Odhner, but it has been much changed and improved. Is hall describe the machine here in some detail, as a typical example. You will find other kinds described in the books mentioned above[3]. My description of course can give you a real understanding of the

[1] Enzyklopädie der mathematischen Wissenschaften, Band I, Teil II. See also *v. Sanden, H.,* Practical Mathematical Analysis (Translation by Levy), Dutton & Co. — *Horsburgh, E. M.,* Modern Instruments and Methods of Calculation. Bell & Sons.

* The *Theory of Exact Calculation.*

** The *Theory of Approximate Calculation.*

[2] Leipzig 1900.

*** *Lectures on Numerical Calculation.*

[[3] Concerning other types of calculating machines, see also A. Galle, *Mathematische Instrumente,* Leipzig 1912.]

machine only if you examine it afterwards personally and if you see, by actual use, how it is operated. The machine will be at your disposal, for that purpose, after the lecture.

So far as the *external appearance* of the Brunsviga is concerned, it presents schematically a picture somewhat as follows (see Fig. 1, p. 18). There is a fixed frame, the *"drum"*, below which and sliding on it, is a smaller longish case, the *"slide"*. A handle which projects from the drum on the right, is operated by hand. On the drum there is a series of parallel slits, each of which carries the digits 0, 1, 2, . . ., 9, read downwards; a peg *s* projects from each slit and can be set at pleasure at any one of the ten digits. Corresponding to each of these slits there is an opening on the slide under which a digit can appear. Figure 3, p. 19 gives a view of a newer model of the machine.

I think that the arrangement of the machine will be clearer if I describe to you the process of carrying out a definite calculation, and the way in which the machine brings it about. For this I select *multiplication*.

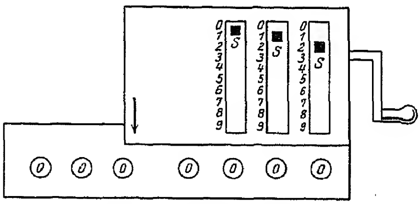


Fig. 1. Before the first turn.

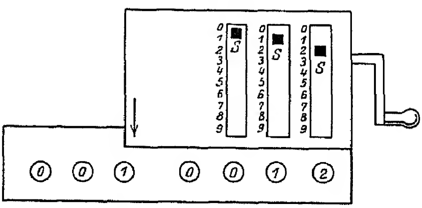


Fig. 2. After the first turn.

The procedure is as follows: One *first sets the drum pegs on the multiplicand*, i. e., beginning at the right, one puts the first lever at the one's digit, the second at the ten's digit of the multiplicand, etc. If, for example, the multiplicand is 12, one sets the first lever at 2, the second lever at 1; all the other levers remain at zero (see Fig. 1). *Now turn the handle once around, clockwise.* The multiplicand appears under the openings of the slide, in our case a 2 in the first opening from the right, a 1 in the second, while zeros remain in all the others. Simultaneously, however, in the first of a series of openings in the slide, at the left, the digit 1 appears to indicate that we have turned the handle once (Fig. 2). *If now one has to do with a multiplier of one digit, one turns the handle as many times as this digit indicates; the multiplier will then be exhibited on the slide to the left, while the product will appear on the slide to the right.* How does the apparatus bring this result about? In the first place there is attached to the under side of the slide, at the left, a cogwheel which carries, equally spaced on its rim, the digits 0, 1, 2, . . ., 9. By means of a driver, this cogwheel is rotated through one tenth of its perimeter with every turn of the handle, so that a digit becomes visible through the opening in the slide, which actually indicates

the number of revolutions, in other words the multiplier. Now as to the *obtaining of the product*, it is brought about by similar cogwheels, one under each opening at the right of the slide. But how is it that by one and the same turning of the handle, one of these wheels, in the above case, moves by one unit, the other by two? This is where the peculiarity in construction of the Brunsviga appears. Under each slit of the drum there is a flat wheel-shaped disc (driver) attached to the axle of the handle, upon which there are nine teeth which are movable in a radial direction (see Fig. 4). By means of the projecting peg S, mentioned above, one can turn a ring R which rests upon the periphery of the disc, so that, according to the mark upon which one sets S in the slit, 0, 1, 2, ..., 9 of the movable teeth spring outward (in Fig. 4, two teeth). These teeth engage the cogs under the corresponding openings of the slide, so that *with one turn of the handle each driver thrusts forward the corresponding cogwheel by as many units as there are teeth pushed out*, i. e., *by as many teeth as one has set with the corresponding peg* S. Accordingly, in the above illustration, when we start at the zero position, and turn the handle once, the units wheel must jump to 2, the ten's wheel to 1, so that 12 appears. A second turn of the handle moves the units wheel another 2 and the tens wheel another 1, so that 24 appears, and similarly, we get, after 3 or 4 times, $3 \cdot 12 = 36$ or $4 \cdot 12 = 48$, respectively.

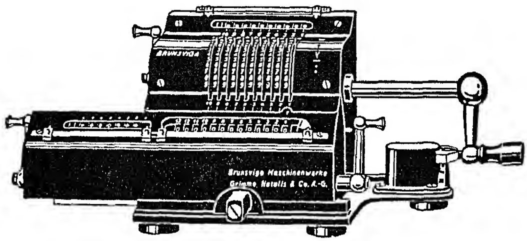


Fig. 3.

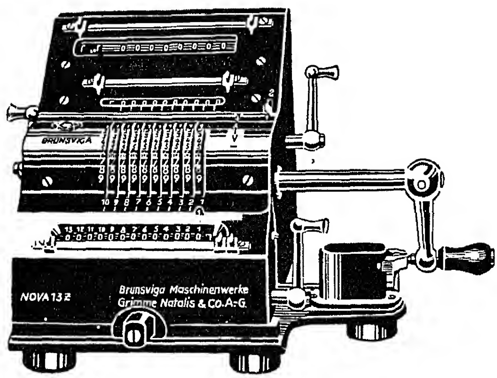


Fig. 3a.

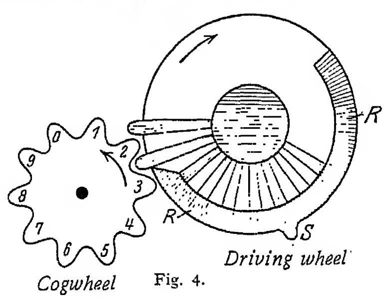


Fig. 4.

But now turn the handle a fifth time: Again, according to the account above, the units wheel should jump again by two units, in other words back to 0, the tens wheel by one, or to 5, and we should have the false result $5 \cdot 12 = 50$. In the actual turning, however, the slide shows 50, to be sure, until just before the completion of the turn; but at the last instant the 5 changes into 6, so that the correct result appears. Something has come into action now that we have not yet described, and which is really the most remarkable point of such machines: the so called *carrying the tens.* Its principle is as follows: *when one of the number bearing cogwheels under the slide* (e. g., *the units wheel*) *goes through zero, it presses an otherwise inoperative tooth of the neighboring driver* (*for the tens*) *into position, so that it engages the corresponding cogwheel* (*the tens wheel*) *and pushes this forward one place farther than it would have gone otherwise.* You can understand the details of this construction only by examining the apparatus itself. There is the less need for my going into particulars here because it is just the method of carrying the tens that is worked out in the greatest variety of ways in the different makes of machines, but I recommend a careful examination of our machine as an example of a most ingenious model. Our collection contains separately the most important parts of the Brunsviga —which are for the most part invisible in the assembled machine—so that you can, by examining them, get a complete picture of its arrangement.

We can best characterize the operation of the machine, so far as we have made its acquaintance, by the words *adding machine, because, with every turn of the handle, it adds, once, to the number on the slide at the right, the number which has been set on the drum.*

Finally, I shall describe in general that arrangement of the machine which permits convenient operation with *multipliers of more than one digit.* If we wish to calculate, say, $15 \cdot 12$ we should have to turn the handle fifteen times, according to the plan already outlined; moreover, if one wished to have the multiplier indicated by the counter at the left of the slide, it would be necessary to have, there also, a device for carrying the tens. Both of these difficulties are avoided by the following arrangement[1]. We first perform the multiplication by five, so that 5 appears on the slide at the left and 60 at the right (see Fig. 5). *Now we push the slide one place to the*

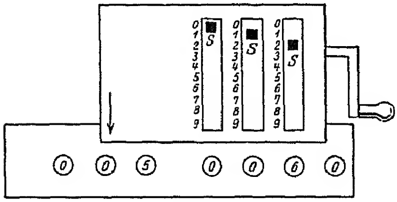


Fig. 5.

[1] In the newer models the cogwheel device for "carrying over" is likewise very complete.

right, so that, as shown in Fig. 5, its units cogwheel is cut out, its tens cogwheel is moved under the units slit of the drum, its hundreds cogwheel under the tens slit, etc., while, at the left, this shift brings it about that the tens cogwheel, instead of the units, is connected with the driver which the handle carries. If we now turn the handle once, 1 appears at the left, in ten's place, so that we read 15; at the right, however, we do not get the addition $\begin{cases} 60 \\ +12 \end{cases}$ but $\begin{cases} \cdot 60 \\ +1\,2\,\cdot \end{cases}$ or, in other words, $60 + 120$, since the 2 is "carried over" to the tens wheel, the 1 to the hundreds wheel. Thus we get correctly $15 \cdot 12 = 180$. It is, as you see, *the exact mechanical translation of the customary process of written multiplication*, in which one writes down under one another, the products of the multiplicand by the successive digits of the multiplier, each product moved to the left one place farther than the preceding, and then adds. *In just the same way one proceeds quite generally when the multiplier has three or more digits, that is, after the usual multiplication by the ones, one moves the slide 1, 2, . . . places to the right and turns the handle in each place as many times as the digit in the tens, hundreds, . . . place of the multiplier indicates.*

Direct examination of the machine will disclose how one can perform other calculations with it; the remark here will suffice that *subtraction and division are effected by turning the handle in the direction opposite to that employed in addition.*

Permit me to summarize by remarking that *the theoretical principle of the machine is quite elementary and represents merely a technical realization of the rules which one always uses in numerical calculation.* That the machine really functions reliably, that all the parts engage one another with unfailing certainty, so that there is no jamming, that the wheels do not turn farther than is necessary, is, of course, the remarkable accomplishment of the man who made the design, and the mechanician who carried it out.

Let us consider for a moment the *general significance of the fact that there really are such calculating machines*, which relieve the mathematician of the purely mechanical work of numerical calculation, and which do this work faster, and, to a higher degree free from error, than he himself could do it, since the errors of human carelessness do not creep into the machine. In the existence of such a machine we see an outright confirmation *that the rules of operation alone, and not the meaning of the numbers themselves, are of importance in calculating;* for it is only these that the machine can follow; it is constructed to do just that; it could not possibly have an intuitive appreciation of the *meaning* of the numbers. We shall not, then, wish to consider it as accidental that such a man as Leibniz, who as both an abstract thinker of first rank and a man of the highest practical gifts, was, at the same tine, both

the father of purely formal mathematics and the inventor of a calculating machine. His machine is, to this day, one of the most prized possessions of the Kästner Museum in Hannover. Although it is not historically authenticated, still I like to assume that when Leibniz invented the calculating machine, he not only followed a useful purpose, but that he also wished to exhibit, clearly, the purely formal character of mathematical calculation.

With the construction of the calculating machine Leibniz certainly did not wish to minimize the *value of mathematical thinking*, and yet it is just such conclusions which are now sometimes drawn from the existence of the calculating machine. If the activity of a science can be supplied by a machine, that science cannot amount to much, so it is said; and hence it deserves a subordinate place. The answer to such arguments, however, is that the mathematician, even when he is himself operating with numbers and formulas, is by no means an inferior counterpart of the errorless machine, "thoughtless thinker" of Thomae; but rather, he sets for himself his problems with definite, interesting, and valuable ends in view, and carries them to solution in appropriate and original manner. He turns over to the machine only certain operations which recur frequently in the same way, and it is precisely the mathematician—one must not forget this—who invented the machine for his own relief, and who, for his own intelligent ends, designates the tasks which it shall perform.

Let me close this chapter with the wish that the calculating machine, in view of its great importance, may become known in wider circles than is now the case. Above all, every teacher of mathematics should become familiar with it, and it ought to be possible to have it demonstrated in secondary instruction.

II. The First Extension of the Notion of Number

With the last section we leave operations with integers, and shall treat, in a new chapter, the *extension of the number concept*. In the schools it is customary, in this field, to take in order the following steps:

1. *Introduction of fractions and operations with fractions.*

2. *Treatment of negative numbers*, in connection with the beginnings of operations with letters.

3. *More or less complete presentation of the notion of irrational numbers by examples that arise upon different occasions*, which leads, then, gradually, to the notion of the *continuum of real numbers*.

It is a matter of indifference in which order we take up the first two points. Let us discuss negative numbers before fractions.

1. Negative Numbers

Let us first note, as to terminology, that in the schools, one speaks of positive and negative numbers, inclusively, as *relative numbers* in distinction from the *absolute* (positive) numbers, whereas, in universities this language is not common. Moreover, in the schools one speaks of "algebraic numbers"[1] along with relative numbers, an expression which we in universities employ, as you know, in quite another sense.

Now, as to the origin and introduction of negative numbers, I can be brief in my reference to source material; these things are already familiar to you, or you can at least easily make them so with the help the references I shall give. You will find a complete treatment, for example, in Weber-Wellstein; also, in very readable form, in H. Burkhardt's Algebraischer Analysis[2]. This book, moreover, you might well purchase, as it is of moderate size.

The creation of negative numbers is motivated, as you know, by the demand *that the operation of subtraction shall be possible in all cases.* If $a < b$ then $a - b$ is meaningless in the domain of natural integers; a number $c = b - a$ does exist, however, and we write

$$a - b = -c$$

which we call a *negative number.* This definition at once justifies the representation of all integers by means of the scale of equidistant points

on a straight line the "axis of abscissas" which extends in both directions from an origin. One may consider this picture as a common possession of all educated persons today, and one can, perhaps, assume that it owes its general dissemination, chiefly, to the thermometer scale. The commercial balance, with its reckoning in debits and credits, affords likewise a graphic and familiar picture of negative numbers.

Let us, however, realize at once and emphatically how extraordinally difficult in principle is the step, which is taken in school when negative numbers are introduced. Where the pupil before was accustomed to represent visually by concrete numbers of things the numbers, and, later, the letters, with which he operated, as well as the results which he obtained by his operations, he finds it now quite different. He has to do with something new, the "negative numbers", which have, immediately, nothing in common with his picture of numbers of things, but he must operate with them as though they had, although the operations

[1] See, e. g. Mehler, *Hauptsätze der Elementarmathematik*, Nineteenth edition, p. 77, Berlin, 1895.

[2] Leipzig 1903. [Third edition, revised by G. Faber, 1920.]—See also Fine, H., *The Number-System of Algebra treated Theoretically and Historically*, Heath.

have graphically a meaning much less clear than the old ones. Here, for the first time, we meet the transition from concrete to formal mathematics. The complete mastery of this transition requires a high order of ability in abstraction.

We shall now inquire in detail what happens to the operations of calculation when negative numbers are introduced. The first thing to notice is that addition and subtraction coalesce, substantially: The addition of a positive number is the subtraction of the equal and opposite negative number. In this connection, Max Simon makes the amusing remark that, whereas negative numbers were created to make the operation of subtraction possible without any exception, subtraction as an independent operation ceased to exist by virtue of that creation. For this new operation of addition (including subtraction) in the domain of positive and negative numbers the five formal laws stated before hold without change. These are, in brief (see p. 9 et seq.):

1. Always possible.
2. Unique.
3. Associative law.
4. Commutative law.
5. Monotonic law.

Notice, in connection with 5, that $a < b$ means, now, that a lies to the left of b in the geometric representation, so that we have, for example $-2 < -1$, $-3 < -2$.

The chief point in the *multiplication* of positive and negative numbers is the *rule of signs*, that $a \cdot (-c) = (-c) \cdot a = -(a \cdot c)$, and $(-c)(-c') = +(c \cdot c')$. Especially the latter rule: "Minus times minus gives plus" is often a dangerous stumbling block. We shall return presently to the inner significance of these rules; just now we shall combine them into a statement defining multiplication of a series of positive and negative numbers: *The absolute value of a product is equal to the product of the absolute values of the factors; its sign is positive or negative according as an even or an odd number of factors is negative.* With this convention, multiplication in the domain of positive and negative numbers has again the following properties:

1. Always possible.
2. Unique.
3. Associative.
4. Commutative.
5. Distributive with respect to addition.

There is a change only in the monotonic law; in its place one has the following law:

6. If $a > b$ then $a \cdot c \gtrless b \cdot c$ according as $c \gtrless 0$.

Let us inquire, now, whether these laws, considered again purely formally, are consistent. We must admit at once, however, that a purely

logical proof of consistency is as yet much less possible here than it is in the case of integers. Only a reduction is possible, in the sense that the present laws are consistent if the laws for integers are consistent. But until this has been completed by a logical consistency proof for integers, one will have to hold that the *consistency of our laws is based solely on the fact that there are intuitive things, with intuitive relations, which obey these laws.* We noted above, as such, the series of integral points on the axis of abscissas and we need only indicate what the rules of operation signify there: The addition $x' = x + a$, where a is fixed, assigns to each point x a second point x', so that the infinite straight line is simply displaced along itself by an amount a, to the right or to the left, according as a is positive or negative. In an analogous manner, the multiplication $x' = a \cdot x$ represents a similarity transformation of the line into itself, a pure stretching for $a > 0$, a stretching together with a reflexion in the origin for $a < 0$.

Permit me now to explain how, historically, all these things arose. One must not think that the negative numbers are the invention of some clever man who menufactured them, together with their consistency perhaps, out of the geometric representation. Rather, during a long period of development, the use of negative numbers forced itself, so to speak, upon mathematicians. Only in the nineteenth century, after men had been operating with them for centuries, was the consideration of their consistency taken up.

Let me preface the history of negative numbers with the remark that the ancient Greeks certainly had no negative numbers, so that one cannot yield them the first place, in this case, as so many people are otherwise prone to do. One must attribute this invention to the Hindus, who also created our system of digits and in particular our zero. In Europe, negative numbers came gradually into use at the time of the Renaissance, just as the transition to operating with letters had been completed. I must not omit to mention here that this completion of operations with letters is said to have been accomplished by Vieta in his book *In Artem Analyticam Isagoge*[1].

From the present point of view, we have the so called parenthesis rules for operations with positive numbers, which are, of course, contained in our fundamental formulas, provided one includes the correrponding laws for subtraction. But I should like to take them up somewhat in detail, by means of two examples, in order, above all, to show the possibility of extremely simple intuitive proofs for them, proofs which need consist only of the representation and of the word "Look"!, as was the custom with the ancient Hindus.

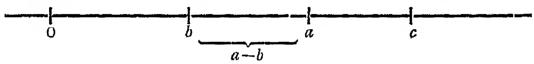
1. *Given* $a > b$ *and* $c > a$, where a, b, c are positive. Then $a - b$ is a positive number and is smaller than c, that is, $c - (a - b)$ must

[1] Tours 1591.

exist as a positive number. Let us represent the numbers on the axis of abscissas and note that the segment between the points b and a has the length $a - b$. A glance at the representation shows that, if we take away from c the segment $a - b$, the result is the same as though we first took away the entire segment a and then restored the part b, i. e.,

$$(1) \qquad c - (a - b) = c - a + b\,.$$

2. *Given* $a > b$ *and* $c > d$; then $a - b$ and $c - d$ are positive integers. We wish to examine the product $(a - b) \cdot (c - d)$; for that purpose



draw the diagonally hatched rectangle (Fig. 6) with sides $a - b$ and $c - d$ whose area is the number sought, $(a - b) \cdot (c - d)$, and which is part of the rectangle with sides a and c. In order to obtain the former rectangle from the latter, we take away first the horizontally hatched rectangle $a \cdot d$, then the vertically hatched one $b \cdot c$; in doing this we have removed twice the double-hatched rectangle $b \cdot d$, and we must put it back. But these operations express precisely the known formula

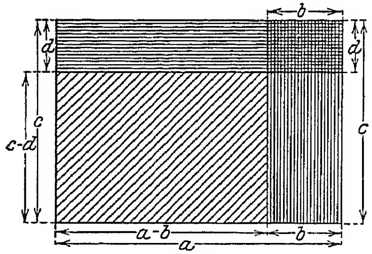


Fig. 6.

$$(2) \quad (a - b)(c - d) = ac - ad - bc + bd\,.$$

As the most important psychological moment to which the introduction of negative numbers, upon this basis of operations with letters, gave rise, that general peculiarity of human nature shows itself, by virtue of which we *are involuntarily inclined to employ rules under circumstances more general than are warranted by the special cases under which the rules were derived and have validity.* This was first claimed as a guiding principle in arithmetic by Hermann Hankel, in his *Theorie der komplexen Zahlsysteme**[1], under the name "Prinzip von der Permanenz der formalen Gesetze"**. I can recommend to your notice this most interesting book. For the particular case before us, of transition to negative numbers, the above principle would declare that one desired to forget, in formulas like (1) and (2) the expressed assumptions as to the relative magnitude of a and b and to employ them in other cases. If one applies (2), for example, to $a = c = 0$, for which the formulas were not proved at all, one obtains $(-b) \cdot (-d) = +bd$, i. e., the *sign rule for multiplication of negative numbers.* In this manner we may derive, in fact almost unconsciously, all the rules, which we must now

* *Theory of Complex Number Systems.*

[1] Leipzig 1867.

** Principle of the permance of formal laws.

designate, following the same line of thought, as *almost necessary assumptions, necessary insofar as one would have validity of the old rules for the new concepts.* To be sure, the old mathematicians were not happy with this abstraction, and their uneasy consciences found expression in names like *invented numbers, false numbers,* etc., which they gave to the negative numbers on occasion. But in spite of all scruples, the negative numbers found more and more general recognition in the sixteenth and seventeenth centuries, because they justified themselves by their usefulness. To this end, the development of analytic geometry without doubt contributed materially. Nevertheless the doubts persisted, and were bound to persist, so long as one continued to seek for a representation in the concept of a number of things, and had not recognized the leading role of formal laws when new concepts are set up. In connection with this stood the continually recurring attempts to prove the rule of signs. The simple explanation, which was brought out in the nineteenth century, is that it is *idle to talk of the logical necessity of the theorem,* in other words, *the rule of signs is not susceptible of proof; one can only be concerned with recognizing the logical permissibility of the rule,* and, at the same time, that it is arbitrary, and regulated by considerations of expedience, such as the principle of permanence.

In this connection one cannot repress that oft recurring thought that things sometimes seem to be more sensible than human beings. Think of it: one of the greatest advances in mathematics, the introduction of negative numbers and of operations with them, was not created by the conscious logical reflection of an individual. On the contrary, its slow organic growth developed as a result of intensive occupation with things, so that it almost seems as though men had learned from the letters. The rational reflection that one devised here something correct, compatible with strict logic, came at a much later time. And, after all, the function of pure logic, when it comes to setting up new concepts, is only to *regulate* and *never to act as the sole guiding principle*; for there will always be, of course, many other conceptual systems which satisfy the single demand of logic, namely, freedom from contradiction.

If you desire still other *literature* concerning questions about the history of negative numbers, let me recommend Tropfkes *Geschichte der Elementarmathematik*[1]*, as an excellent collection of material containing, in lucid presentation, a great many details about the development of elementary notions, views, and names.

---

[1] Two volumes, Leipzig 1902/03. [Second edition revised and much enlarged, to appear in seven volumes, of which six had appeared by 1924.]—See also Cajori, F., *History of Mathematics*, Macmillan.

* *History of Elementary Mathematics.*

If we now look critically at the way in which negative numbers are presented in the schools, we find frequently the error of trying to prove the logical necessity of the rule of signs, corresponding to the above noted efforts of the older mathematicians. One is to derive $(-b)(-d) = +bd$ heuristically, from the formula $(a - b)(c - d)$ and to think that one has a proof, completely ignoring the fact that the validity of this formula depends on the inequalities $a > b$, $c > d$[1]. Thus the proof is fraudulent, and the psychological consideration which would lead us to the rule by way of the principle of permanence is lost in favor of quasi-logical considerations. Of course the pupil, to whom it is thus presented for the first time, cannot possibly comprehend it, but in the end he must nevertheless believe it; and if, as it often happens, the repetition in a higher class does not supply the corrective, the conviction may become lodged with some students that the whole thing is mysterious, incomprehensible.

In opposition to this practice, I should like to urge you, in general, never to attempt to make impossible proofs appear valid. One should convince the pupil by simple examples, or, if possible, let him find out for himself that, in view of the actual situation, *precisely these conventions, suggested by the principle of permanence, are appropriate in that they yield a uniformly convenient algorithm, whereas every other convention would always compel the consideration of numerous special cases.* To be sure, one must not be precipitate, but must allow the pupil time for the revolution in his thinking which this knowledge will provoke. And while it is easy to understand that other conventions are not advantageous, one must emphasize to the pupil how really wonderful the fact is that a general useful convention really *exists*; it should become clear to him that this is by no means self-evident.

With this I close my discussion of the theory of negative numbers and invite you now to give similar consideration to the second extension of the notion of number.

2. Fractions.

Let us begin with the treatment of fractions in the schools. There the fraction a/b has a thoroughly concrete meaning from the start. In contrast to the graphic picture of the integer, there has been only a change of base: We have passed from the *number* of things to their *measure*, from the consideration of *countable things* to *measurable things.* The system of *coins*, or of *weights*, affords, with some restriction, and the system of *lengths* affords completely, an example of *measurable manifolds*. These are the examples with which the idea of the fraction is

[1] See, for example, E. Heis, *Sammlung von Beispielen und Aufgaben aus der Arithmetik und Algebra.* Edition 1904, p. 46, 106—108.

given to every pupil. No one has great difficulty in grasping the meaning of $^1/_3$ meter oder $^1/_2$ pound. The *relations* $=, >, <$, between fractions can be immediately developed by means of the same concrete intuition, and likewise the operations of *addition* and *subtraction*, as well as the multiplication of a fraction *by an integer*. After this, *general multiplication* can easily be made comprehensible: *To multiply a number by a/b means to multiply it by a and then to divide by b; in other words: the product is derived from the multiplicand just as a/b is derived from* 1. *Division* by a fraction is then presented as the *operation inverse* to multiplication: *a divided by $2/3$ is the number which multiplied by $2/3$ gives a.* These notions of operations with fractions combine with that of negative numbers so that one finally has the totality of all rational numbers. I cannot enter into the details of this building-up process, which, in the school, takes, of course, a long time. Let us rather compare it at once with the perfected presentation of modern mathematics, using for this purpose the above mentioned books of Weber-Wellstein and Burkhardt[1].

Weber-Wellstein emphasizes primarily the formal point of view which, from the multiplicity of possible interpretations, selects what is of necessity common to all. According to this view, the fraction a/b is a symbol, a "number-pair" with which one can operate according to certain rules. These rules, which in our discussion above arose naturally from the meaning of fraction, have here the character of arbitrary conventions. For example, that which, to the pupil, is an obvious theorem concerning the multiplication or division of both terms of a fraction by the same number, appears here as a definition of equality: *two fractions a/b, c/d are called equal when $ad = bc$.* Similarly, *greater than* and *smaller than* are defined, and one *agrees* that *the fraction* $\left(\frac{ad + bc}{bd}\right)$ *shall be called the sum of the two fractions a/b, c/d, etc.* It is thus *proved* that the operations, so defined in the new domain of numbers, possess formally exactly the properties of addition and multiplication for integers, i. e., they satisfy the eleven fundamental laws which have been repeatedly enumerated.

Burkhardt does not proceed quite so formally as does Weber-Wellstein, whose presentation we have sketched in its essentials. He looks upon the *fraction a/b as a sequence of two operations in the domain of integers*: *a multiplication by a and a division by b*, in which the object upon which these operations are performed is an arbitrarily chosen integer. If one undertakes two such *"pairs of operations"* a/b, c/d, this is said to correspond to *multiplication of the fractions*, and one sees easily that the operation so resulting is none other than multiplication by $a \cdot c$ and division by $b \cdot d$, so that the *rule for the multiplication of fractions*,

[1] In what follows, the first editions of these books have been used.

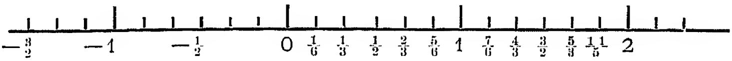
$\left(\frac{a}{b}\right) \cdot \left(\frac{c}{d}\right) = \left(\frac{a \cdot c}{b \cdot d}\right)$, *is obtained out of the clear meaning of the fractions*, but not determined merely as an arbitrary convention. One can, of course, treat division in the same way. Addition and subtraction, on the other hand, do not admit of such a simple explanation with this representation; thus the formula $\frac{a}{b} + \frac{c}{d} = \frac{(a\,d + b\,c)}{b\,d}$ remains, with Burkhardt also, only a convention for which he adduces only reasons of plausibility.

Let us now compare the older presentation in the schools, with the modern conception just sketched. According to the latter, in the one book as well as in the other, we are left really completely in the field of integers, in spite of the extension of the notion of number. It is merely assumed that the totality of whole numbers is intuitively grasped, or that the rules of operation with them are known; the things newly defined as number-pairs, or as operations with whole numbers, fit completely into this frame. The school treatment, on the other hand, is based entirely on the newly acquired conception of measurable quantities, which supplies an immediate intuitive picture of fractions. We can best grasp this difference if we imagine a being who has the notion of whole numbers, but no conception of measurable quantities. For him the school presentation would be wholly unintelligible, whereas he could well comprehend the discussions of either Weber-Wellstein or Burkhardt.

Which of the two methods is the better? What does each accomplish? The answer to this will be like the one we gave recently when we put the analogous question concerning the different conceptions of integers. The modern presentation is surely purer, but it is also less rich. For, of that which the traditional curriculum supplies as a unit, it gives really only one part: the abstract and logically complete introduction of certain arithmetic concepts, called "*fractions*", and of operations with them. But it leaves unexplained an entirely independent and no less important question: Can one really apply the theoretical doctrine so derived, to the concrete measurable quantities about us? Again one could call this a problem of "applied mathematics", which admits an entirely independent treatment. To be sure, it is questionable whether such a separation would be desirable pedagogically. In Weber-Wellstein, moreover, this splitting of the problem into two parts finds characteristic expression. After the abstract introduction of operations with fractions, of which alone we have thus far taken account, they devote a special (the fifth) division—called "ratios"—to the question of applying *rational numbers to the external world*. The presentation is, to be sure, rather abstract than intuitive.

I shall now close this discussion of fractions with a general remark concerning the totality of rational numbers, where, for the sake of clearness, I shall make use of the representation upon a straight line.

Think of all points with rational abscissas marked upon this line; we designate them briefly as rational points. We say, then, that the totality of these rational points on the axis of abscissas is "dense", meaning that in every interval, however small, there are still infinitely many



rational points. If we wish to avoid putting anything new into the notion of rational numbers, we might say, more abstractly, that between any two rational points there is always another rational point. It follows that one can separate from the totality of rational points, finite parts which contain neither a smallest nor a largest element. The totality of all rational points between 0 and 1, these points excluded, is an example. For, given any number between 0 and 1, there would still be a number between it and 0, i. e., a smaller, and a number between it and 1, i. e., a larger. In their systematic development, these concepts belong to the *theory of point sets* of Cantor. In fact, we shall make use later of the totality of rational numbers, together with the property just mentioned, as an important example of a point set.

I shall pass now to the third extension of the number system: the irrational numbers.

3. Irrational Numbers.

Let us not spend any time in discussing how this field is usually treated in the schools, for there one does not get much beyond a few examples. Let us rather proceed at once to the historical development. Historically, the origin of the concept of irrational numbers lies certainly in geometric intuition and in the requirements of geometry. If we consider, as we did just now, that the set of rational points is *dense* on the axis of abscissas, then there are still other points on it. Pythagoras is said to have shown this in a manner somewhat as follows. Given a right triangle with each leg of length 1, then the hypotenuse is of length $\sqrt{2}$, and this is certainly not a rational number; for if one puts $\sqrt{2} = \frac{a}{b}$ where a and b are integers, prime to each other, one is led easily by the laws of divisibility of integers to a contradiction. *If we now lay off geometrically on the axis of abscissas, beginning at zero, the segment thus constructed, we obtain a non-rational point which is not one of the original set that is dense on the axis.* Furthermore, the Pythagoreans certainly were aware that, in most cases, the hypotenuse, $\sqrt{m^2 + n^2}$, of a right triangle with legs m and n, is irrational. The discovery of this extraordinarily essential fact was indeed worth the sacrifice of one hundred

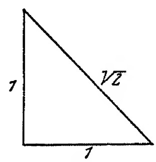


Fig. 7.

oxen with which Pythagoras is said to have celebrated it. We k that the Pythagorean School was fond of searching out those spe of values for m and n for which the right triangle has *three comm sides*, whose lengths, in an appropriately selected unit of mea be expressed in integers (so called Pythagorean numbers). The example of one of these number-triples is $3, 4, 5$.

Later Greek mathematicians studied, in addition to these irrationalities, others that were more complicated; thus one Euclid types such as $\sqrt{\sqrt{a}+\sqrt{b}}$, and the like. We may say, in general, that they confined themselves essentially to such irı ties as one obtains by repeated extraction of square root, ar can therefore be constructed geometrically with ruler and co The general idea of irrational number was not yet known 1

I must, modify this remark somewhat, however, in order misunderstanding. The more precise statement is that the possessed no method for producing or defining, arithmetic general irrational number in terms of rational numbers. This is of modern development and will soon engage our attention. Neve from another point of view they were familiar with the notio general real number which was not necessarily rational; but the had an entirely different appearance to them because they did letters for general numbers. In fact they studied, and Euclid d very systematically, *ratios of two arbitrary segments*. They opera such ratios precisely as we do today with arbitrary real numb deed we find in Euclid definitions which suggest strongly the theory of irrational numbers. Moreover the name used is differ that of the natural number; the latter is called *ἀριθμος*, whe line ratio, the arbitrary real number, is called *λόγος*.

I should like to add a remark concerning the word "*irratio* is without doubt the translation into Latin of the Greek "(The Greek word, however, meant presumably "inexpressible" plied that the new numbers, or line ratios, could not, like the numbers, be expressed by the ratio of two whole numbers misunderstanding put upon the Latin "ratio", that it could only the meaning "reason", gave to "irrational" the meaning "ι nable", which seems still to cling to the term *irrational num*

The general idea of the irrational number appeared first at of the sixteenth century as a consequence of the introduction of fractions, the use of which became established at that time in con with the appearance of logarithmic tables. If we transform a ı number into a decimal, we may obtain *infinite decimals*[2], as well

---

[1] See Tropfke, second edition, Vol. 2, p. 71.

[2] For complete treatment of this subject see, p. 40 et seq.

decimals, but they will always be *periodic*. The simplest example is $\frac{1}{3} = 0.333\ldots$, i.e., a decimal whose period of one digit begins immediately after the decimal point. Now there is nothing to prevent our thinking of an aperiodic decimal whose digits proceed according to any definite law whatever, and anyone would instinctively consider it as a definite, and hence a non-rational, number. By this means the general notion of irrational number is established. It arose to a certain extent automatically, by the consideration of decimal fractions. Thus, historically, the same thing happened with irrational numbers that, as we have seen, happened with negative numbers. Calculation forced the introduction of the new concepts, and without being concerned much as to their nature or their motivation, one operated with them, the more particularly since they often proved to be extremely useful.

It was not until the sixth decade of the nineteenth century that the need was felt for a more precise arithmetic formulation of the foundations of irrational numbers. This occurred in the lectures which Weierstrass delivered at about that date. In 1872, a general foundation was laid simultaneously by G. Cantor of Halle, the founder of the theory of point sets, and independently by R. Dedekind of Braunschweig. I will explain Dedekind's point of view in a few words. Let us assume a knowledge of the totality of rational numbers, but let us exclude all space perception, which would force upon us forthwith the notion of the continuity of the number series. With this understanding, in order to attain to a purely arithmetic definition of the irrational number, Dedekind sets up the notion of a "*cut*" in the domain of rational numbers. If r is any rational number, it separates the totality of rational numbers into two parts A and B *such that every number in A is smaller than any number in B and every rational number belongs to one of these two classes.* A is the totality of all rational numbers which are smaller than r, B those that are larger, whereby r itself may be thought of indifferently as belonging to the one or to the other. Besides these "*proper cuts*" there are also "*improper cuts*", these being separations of all rational numbers into two classes having the same properties except that they are not brought about by a rational number, i. e., separations such that there is neither a smallest rational number in B nor a largest in A. An example of such an improper cut is supplied by, say, $\sqrt{2} = 1.414\ldots$ In fact, every infinite decimal fraction defines a cut, provided one assigns to B every rational number which is larger than every approximation to the infinite decimal, and to A every other rational number; each number in A would thus be equalled or exceeded by at least one approximation (and hence by infinitely many). One can easily show that this cut is proper if the decimal is periodic, improper if it is not periodic.

With these considerations as his basis, Dedekind sets up his definition, which, from a purely logical standpoint, must be looked upon as an

arbitrary convention: *A cut in the domain of rational numbers is called a rational number or an irrational number according as the cut is proper or improper.* A definition of equality follows from this at once: *Two numbers are said to be equal if they yield the same cut in the domain of rational numbers.* From this definition we can immediately *prove* for example, that, $^1/_3$ is equal to the infinite decimal 0.3333 .... If we accept this standpoint, we must demand a proof, i. e., a process of reasoning depending upon the definition given, although this would appear quite unnecessary to one approaching the subject naively. Moreover, such a proof is immediate, if one reflects that every rational number smaller than $^1/_3$ will be exceeded ultimately by the decimal approximations, whereas these are smaller than every rational number which exceeds $\frac{1}{3}$. The corresponding definition in the lectures of Weierstrass appears in the following form: *Two numbers are called equal if they differ by less than any preassigned constant, however small.* The connection with the preceding explanation is clear. The last definition becomes striking if one reflects why 0.999 ... is equal to 1; the difference is certainly smaller than 0.1, smaller than 0.01, etc., that is, it is exactly zero, according to the definition.

If we enquire how it happens that we can admit the irrational numbers into the system of ordinary numbers and operate with them in just the same way, the answer is to be found in the validity of the monotonic law for the four fundamental operations. The principle is as follows: *If we wish to perform upon irrational numbers the operation of addition, multiplication, etc., we can enclose them between ever narrowing rational limits and perform upon these limits the desired operations; then, because of the validity of the monotonic law, the result will also be enclosed between ever narrowing limits.*

It is hardly necessary for me to explain these things in greater detail, since very readable presentations of them are easily available in many books, especially in Weber-Wellstein and in Burkhardt. I hope that you will read more fully than I could tell you here in these books, about the definition of irrational numbers.

I should prefer, rather, to talk about something which you will hardly find in the books, namely, how, after establishing this arithmetic theory, we can pass to the applications in other fields. This applies in particular, to analytic geometry, which to the naive perception appears to be (and psychologically really is) the source of irrational numbers. If we think of the axis of abscissa, with the origin and also the rational points marked on it, as above, then these applications depend upon the following fundamental principle: *Corresponding to every rational or irrational number there is a point which has this number as abscissa and, conversely, corresponding to every point on the line there is a rational or an irrational number, viz., its abscissa.* Such a fundamental principle,

which stands at the head of a branch of knowledge, and from which all that follows is logically deduced, while it itself cannot be logically proved, may properly be called an *axiom*. Such an axiom will appear intuitively obvious or will be accepted as a more or less arbitrary convention, by each person according to his gifts. This axiom concerning the one-to-one correspondence between real numbers on one hand, and the points of a straight line on the other, is usually called the *Cantor axiom* because *G. Cantor* was the first to formulate it specifically (in the Mathematische Annalen, vol. 5, 1872).

This is the proper place to say a word about the nature of space perception. It is variously ascribed to two different sources of knowledge. One the sensibly immediate, the empirical intuition of space, which we can control by means of measurement. The other is quite different, and consists in a subjective idealizing intuition, one might say, perhaps, our *inherent idea of space*, which goes beyond the inexactness of sense observation. I pointed out to you an analogous difference when we were discussing the notion of number. We may characterize it best as follows: It is immediately clear to us what a small number means, like 2 or 5, or even 7, whereas we do not have such immediate intuition of a larger number, say 2503. Immediate intuition is replaced here by the subjective intuition of an ordered number series, which we derive from the first numbers by mathematical induction. There is a similar situation regarding space perception. Thus, if we think of the distance between two points, we can estimate or measure it only to a *limited degree of exactness*, because our eyes cannot recognize as different two line-segments whose difference in length lies below a certain limit. This is the concept of the *threshold* of perception which plays such an important role in psychology. This phenomenon still persists, in its essentials, when we aid the eye with instruments of the highest precision; for there are physical properties which prohibit our exceeding a certain degree of exactness. For instance, optics teaches that the wave-length of light, which varies with the color, is of the order of smallness of $^1/_{1000}$ mm. (= 1 micron); it shows also that objects whose dimensions are of this order of smallness cannot be seen distinctly with the best microscopes because diffraction enters then and hence no optical image can give exact reproductions of the details. *The result of this is the impossibility, by direct optical means, of getting measures of length that are finer than to within one micron, so that, when measured lengths are given in millimeters, only the first three decimals can have an assured meaning.* In the same way, in all physical observations and measurements, one meets such threshold values which cannot be passed, which determine the extreme limits of possible exactness of lengths which have been measured and expressed in millimeters. Statements beyond this limit have no meaning, and are an evidence of ignorance or of attempted deception. One often

finds such excessively exact numbers in the advertisements of medicinal springs, where the percentage of salt, which really varies with the time, is given to a number of decimal places which could not possibly be determind by weighing.

In contrast with this property of empirical space perception which is restricted by limitations on exactness, *abstract, or ideal space perception demands unlimited exactness, by virtue of which, in view of Cantor's axiom, it corresponds exactly to the arithmetic definition of the number concept.*

In harmony with this division of our perception, it is natural to divide mathematics also into two parts, which have been called *mathematics of approximation* and the *mathematics of precision*. If we desire to explain this difference by an interpretation of the equation $f(x) = 0$, we may note that, in the mathematics of approximation, just as in our empirical space perception, one is not concerned that $f(x)$ should be *exactly* zero, but merely that its absolute value $|f(x)|$ should *remain below the attainable threshold of exactness* ε. The symbol $f(x) = 0$ is merely an abbreviation for the inequality $|f(x)| < \varepsilon$, with which one is really concerned. It is only in the mathematics of precision that one insists that the equation $f(x) = 0$ be exactly satisfied. Since mathematics of approximation alone plays a rôle in applications, one might say, somewhat crassly, that one needs only this branch of mathematics, whereas the mathematics of precision exists only for the intellectual pleasure of those who busy themselves with it, and to give valuable and indeed indispensable support for the development of mathematics of approximation.

In order to return to our real subject, I add here the remark that the *concept of irrational number belongs certainly only to mathematics of precision*. For, the assertion that two points are separated by an irrational number of millimeters cannot possibly have a meaning, since, as we saw, when our rigid scales are measured in meters, all decimal places beyond the sixth are devoid of meaning. *Thus in practice we can, without concern, replace irrational numbers by rational ones.* This may seem, to be sure, to be contradicted by the fact that, in crystallography, one talks of the law of rational indices, or by the fact that in astronomy, one distinguishes different cases according as the periods of revolution of two planets have a rational or an irrational ratio. In reality, however, this form of expression only exhibits the many-sidedness of language; for one is using here rational and irrational in a sense entirely different from that hitherto used, namely, in the sense of *mathematics of approximation*. In this sense, one says that two magnitudes have a rational ratio when they are to each other as two *small* integers, say $3/7$; whereas one would call the ratio $2021/7053$ irrational. We cannot say how large numerator and denominator in this second case must be, in general, since that depends upon the problem in hand. I discussed all these

interesting relations in a course of lectures in the Summer Semester of 1901, which was lithographed in 1902 and which will constitute the third volume of the present work (see the preface to the third edition, p. V): *Applications of Differential and Integral Calculus to Geometry, a Revision of Principles* [Elaborated by C. H. Müller].

In conclusion let me say, in a few words, how I would have these matters handled in the schools. An exact theory of irrational numbers would hardly be adapted either to the interest or to the power of comprehension of most of the pupils. The pupil will usually be content with results of limited exactness. He will look with astonished approval upon correctness to within $^1/_{1000}$ mm and will not demand unlimited exactness. For the average pupil it will be sufficient if one makes the irrational number intelligible in general by means of examples, and this is what is usually done. To be sure, especially gifted individual pupils will demand a more complete explanation than this, and it will be a laudable exercise of pedagogical skill on the part of the teacher to give such students the desired supplementary explanation without sacrificing the interests of the majority.

III. Concerning Special Properties of Integers

We shall now begin a new chapter which will be devoted to the *actual theory of integers*, to the *theory of numbers*, or *arithmetic in its narrower sense*. I shall first recall in tabular form the individual questions from this science which appear in the school curriculum.

1. The first problem of the theory of numbers is that of *divisibility*: Is one number divisible by another or not?

2. Simple rules can be given which enable us easily to decide as to the *divisibility of any given number by smaller numbers*, such as 2, 3, 4, 5, 9, 11, etc.

3. There are *infinitely many prime numbers*, that is, numbers which have no *integral divisors* except one and themselves): 2, 3, 4, 5, 9, 11, etc.

4. We are in control of all of the properties of given integers if we know their decomposition *into prime factors*.

5. In the *transformation of rational fractions into decimal fractions* the theory of numbers plays an important role; it shows why the decimal fraction must be *periodic* and how large the period is.

Although such questions may be considered in secondary schools, when the pupils are between the ages of eleven and thirteen, the theory of numbers comes up only in isolated places during the later years, and, at most, the following points are considered.

6. *Continued fractions* are taught occasionally, although not in all schools.

7. Sometimes instruction is given also in Diophantine equations, that is, equations with several unknowns which can take only integral values.

The *Pythagorean numbers* of which we spoke (see p. 32), furnish an example; here one has to do with triplets of integers which satisfy the equation

$$a^2 + b^2 = c^2.$$

8. The *problem of dividing the circle into equal parts* is closely related to the theory of numbers, although the connection is hardly ever worked out in the schools. If we wish to divide the circle into n equal parts, using, of course, *only ruler and compasses*, it is easy to do it for $n = 2, 3, 4, 5, 6$. It cannot be done, however, if $n = 7$, hence we stop respectfully when we come to this problem in the school. To be sure, it is not always stated definitely that this construction is really impossible when $n = 7$,—a fact whose explanation lies somewhat deep in number-theoretic considerations. In order to forestall misunderstandings, which unfortunately often arise, let me say, with emphasis, that one is concerned here again with a *problem of mathematics of precision*, which is devoid of meaning for the applications. In practice, even in cases where an "exact" construction is possible, it would not be used ordinarily; for, in the field of mathematics of approximation, the circle can be divided into any desired number of equal parts more suitably by simple skillful experiment; and any prescribed, practically possible, degree of exactness can be attained. Every mechanician who makes instruments that carry divided circles proceeds in this way.

9. The higher theory of numbers is touched by the school curriculum in one other place, namely, when *π is calculated*, during the study of the *quadrature* of the circle. We usually determine the first decimal places for π, by some method or other, and we mention incidentally, perhaps, the *modern proof of the transcendence of π which sets at rest the old problem of the quadrature of the circle with ruler and compasses*. At the end of this course I shall consider this proof in detail. For the present I shall give merely a prescise formulation of the fact, namely, *that the number π does not satisfy any algebraic equation with integral coefficients*:

$$a\pi^n + b\pi^{n-1} + \cdots + k\pi + 1 = 0.$$

It is especially important that the coefficients be integers, and it is for this reason that the problem belongs to the theory of numbers. Of course here, again, one is concerned solely with a *problem of the mathematics of precision*, because it is only in this sense that the number-theoretic character of π has any significance. The mathematics of approximation is satisfied with the determination of the first few decimals, which permit us to effect the quadrature of the circle with any desired degree of exactness.

I have sketched for you the place of the theory of numbers in the schools. Let us consider now its proper place in *university instruction* and in *scientific investigation*. In this connection I should like to divide

research mathematicians, according to their attitude toward theory of numbers, into two classes, which I might call the *enthusiastic* class and the *indifferent* class. For the former there is no other science so beautiful and so important, none which contains such clear and precise proofs, theorems of such impeccable rigor, as the theory of numbers. Gauss said "If mathematics is the queen of sciences, then the theory of numbers is the queen of mathematics". On the other hand, theory of numbers lies remote from those who are indifferent; they show little interest in its development, indeed they positively avoid it. The majority of students might, as regards their attitude, be put into the second class.

I think that the *reason for this remarkable division* can be summarized as follows: On the one hand the theory of numbers is *fundamental for all more thoroughgoing mathematical research*; proceeding from entirely different fields, one comes at last, with extraordinary frequency, upon relatively simple arithmetic facts. On the other hand, however, the *pure theory of numbers is an extremely abstract thing*, and one does not often find the gift of ability to understand with pleasure anything so abstract. The fact that most textbooks are at pains to present the subject in the most abstract way tends to accentuate this unattractiveness of the subject. I believe that *the theory of numbers would be made more accessible, and would awaken more general interest, if it were presented in connection with graphical elements and appropriate figures*. Although its theorems are logically independent of such aids, still one's comprehension would be helped by them. I attempted to do this in my lectures in 1895/96[1] and a similar plan is followed by H. Minkowski in his book on *Diophantische Approximationen*[2]. My lectures were of a more elementary introductory character, whereas Minkowski considers at an early point special problems in a detailed manner.

As to *textbooks in the theory of numbers*, you will often find all you need in the textbooks in algebra. Among the large number of books on the theory of real numbers, I would mention especially Bachman's *Grundlagen der neueren Zahlentheorie*[3].

In the *more special number-theoretic discussions* which I shall give here, I shall keep touch with the points mentioned above and I shall endeavor especially to present the matter as graphically as possible. While I shall restrict myself to material that is *valuable for the teacher*, I shall by no means put it into a form suitable for immediate presentation to the pupils. The necessity for this arises from my *experiences in*

---

[1] *Ausgewähltes Kapitel der Zahlentheorie* (mimeographed lectures written up by A. Sommerfeld and Ph. Furtwängler). Second printing (already exhausted). Leipzig 1907.

[2] With an appendix: *Eine Einführung in die Zahlentheorie*. Leipzig 1907.

[3] Sammlung Schubert No. 53. Leipzig 1907. [Second edition published by R. Hauszner 1921.] — See also Carmichael, R. D., *Theory of Numbers*. Wiley.

examinations, which show me that the number-theoretic information of candidates is often confined to catchwords which have no thorough knowledge back of them. Every candidate can tell me that π is "transcendental"; but many of them do not know what that means; I was told, once, that a transcendental number was neither rational nor irrational. Likewise I often find candidates who tell me that the number of primes is infinite, but who have no notion as to the proof, although it is so simple.

I shall start my number-theoretic discussion with this proof, assuming that you are acquainted with the first two points metioned in our list. As a matter of history I remind you that this proof was handed on to us by *Euclid*, whose "elements" (Greek στοιχεῖα) contained not only his system of geometry, but also *algebraic and arithmetic information in geometric language*. Euclid's transmitted *proof of the existence of infinitely many prime numbers* is as follows: Assuming that the sequence of prime numbers is finite, let it be $1, 2, 3, 5, \ldots, p$; then the number $N = (1 \cdot 2 \cdot 3 \cdot 5 \ldots p) + 1$ is not divisible by any of the numbers $2, 3, 5, \ldots p$ since there is always the remainder 1; hence N must either itself be a prime number or there are prime numbers larger than p. Either of these alternatives contradicts the hypothesis, and the proof is complete.

In connection with the *fourth point*, the *separation into prime factors*, I should like to call to your attention one of the older *factor tables*: *Chernac, Cribum Arithmeticum*[1], a large, meritorious work which deserves, historically, all the more attention because it is so reliable. The name of the table suggests the sieve of *Eratosthenes*. The idea on which it was based is that we should discard gradually from the series of all integers those which are divisible by $2, 3, 5, \ldots$, so that only the prime numbers would remain. Chernac gives the decomposition into prime factors of all integers up to 1020000 which are not divisible by 2, 3, or 5; all the prime numbers are marked with a bar. It was in the Chernac work that all the prime numbers lying within the limits stated above were first given. During the nineteenth century the determination was extended to all prime numbers as far as nine million.

I turn now to the *fifth point*, the transformation *of ordinary fractions into decimal fractions*. For the complete theory I shall refer you to Weber-Wellstein, and I shall explain here only the principle of the method by means of a typical example. Let us consider the fraction $1/p$, where p is a prime number different from 2 and 5. We shall show *that $1/p$ is equal to an infinite periodic decimal, and that the number δ of places in the period is the smallest exponent for which 10^δ, when divided by p, leaves* 1

[1] Deventer 1811.

as a remainder, or that, in the language of number theory, δ *is the smallest exponent which satisfies the "congruence"*:

$$10^\delta \equiv 1 \pmod{p}.$$

The proof requires, in the first place, the knowledge that this congruence always has a solution. This is supplied by the *theorem of Fermat*, which states that for every prime number p except 2 and 5:

$$10^{p-1} \equiv 1 \pmod{p}.$$

We shall omit here the proof of this fundamental theorem, which is one of the permanent tools of every mathematician. Secondly, we must borrow from the theory of numbers the theorem that the smallest exponent in question, δ, is *either* $p - 1$ *itself or a divisor of* $p - 1$. We can apply this to the given value p and find that $\frac{10^\delta - 1}{p}$ is an integer N so that one has:

$$\frac{10^\delta}{p} = \frac{1}{p} + N.$$

If we now think of $10^\delta/p$, as well as $1/p$, converted into a decimal, the digits in the two decimals must be identical, since the difference is an integer. But since $10^\delta/p$ is got from $1/p$ by moving the decimal point δ places to the right, it follows that the digits in the decimal expression of $1/p$ are unaltered by this operation, in other words *that the decimal fraction* $1/p$ *consists of continued repetition of the same "period" of* δ *digits*.

In order now to see that there *cannot be a smaller period of* $\delta' < \delta$ *digits* one needs only to prove that the digit number δ' of *every* period must satisfy the congruence $10^{\delta'} \equiv 1$; for we know that δ was the *smallest* solution of this congruence. This proof will result if we pursue the preceding argument in the reverse direction. It follows from our assumption that $1/p$ and $10^{\delta'}/p$ coincide in their decimal places, hence that $\frac{10^{\delta'}}{p} - \frac{1}{p}$ is an integer N', and therefore that $10^{\delta'} - 1$ is divisible by p, or, in other words, that $10^{\delta'} \equiv 1 \pmod{p}$. This completes the proof.

I will give you a few of the simplest instructive *examples*, which will show that δ can take widely different values, both smaller than and equal to $p - 1$. Notice first that for:

$$\tfrac{1}{3} = 0.333\ldots$$

the number of digits in the period is 1, and that in fact, $10^1 \equiv 1 \pmod{3}$. Similarly we find

$$\tfrac{1}{11} = 0.0909\ldots,$$

whence $\delta = 2$, and correspondingly $10^1 \equiv 10, 10^2 \equiv 1 \pmod{11}$. The maximum value $= p - 1$ appears in the example:

$$\tfrac{1}{7} = 0.142857142857\ldots.$$

Here $\delta = 6$ and we have, in fact, $10^1 \equiv 3$, $10^2 \equiv 2$, $10^3 \equiv 6$, $10^4 \equiv 4$, $10^5 \equiv 5$, and $10^6 \equiv 1 \pmod 7$.

Now let us take up, in a similar way, the *sixth point* of my list, *continued fractions*. I shall not present this, however, in the usual abstract arithmetic manner, since you will find it given elsewhere, e. g., in Weber-Wellstein. I shall take this opportunity to show you how number-theoretic things take on a clear and easily intelligible form through geometric and graphical presentation. In this use of geometric aids in number theory we are really only retracing the steps followed by *Gauss* and *Dirichlet*. It was the later mathematicians, say from 1860 on, who banished geometric methods from the theory of numbers. Of course, I can give here only the most important trains of thought and theorems, without proof, and I shall assume that you are not entire strangers to the elementary theory of continued fractions. My lithographed lectures on number theory[1] contain a thoroughgoing account.

You know how the *development of a given positive number* ω *into a continued fraction* arises. We separate out the largest positive integer n_0 contained in ω and write:

$$\omega = n_0 + r_0, \qquad \text{where } 0 \leqq r_0 < 1,$$

then, if $r_0 \neq 0$, we treat $1/r_0$ as we did ω:

$$1/r_0 = n_1 + r_1, \qquad \text{where } 0 \leqq r_1 < 1,$$

and continue in the same way:

$$1/r_1 = n_2 + r_2, \qquad \text{where } 0 \leqq r_2 < 1,$$
$$1/r_2 = n_3 + r_3, \qquad \text{where } 0 \leqq r_3 < 1,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

The process *terminates after a finite number of steps if* ω *is rational*, because a vanishing remainder r_ν must appear in that case; otherwise the process goes on indefinitely. In any case, we write, as the *development of* ω *into a continued fraction*:

$$\omega = n_0 + \cfrac{1}{n_1 + \cfrac{1}{n_2 + \cfrac{1}{n_3 + \cdots}}}$$

As an example, the continued fraction for π is

$$\pi = 3{\cdot}14159265 \cdots = 3 + \cfrac{1}{7 + \cfrac{1}{15 + \cfrac{1}{1 + \cfrac{1}{292 + \cdots}}}},$$

[1] See also Klein, F., *Gesammelte Mathematische Abhandlungen*, Vol. II, pp. 209 to 211.

If we stop the development after the first, second, third, . . . partial denominator, we obtain rational fractions, called *convergents*:

$$n_0 = \frac{p_0}{q_0}, \qquad n_0 + \frac{1}{n_1} = \frac{p_1}{q_1}, \qquad n_0 + \cfrac{1}{n_1 + \cfrac{1}{n_2}} = \frac{p_2}{q_2}, \ \ldots;$$

these give remarkably good approximations to the number ω, or, to speak more exactly, each one of them gives an *approximation which is closer than that given by any other rational fraction which does not have a larger denominator.* Because of this property, continued fractions are of practical importance where one seeks the best possible approximation to an irrational number, or to a fraction with a large denominator (e. g. a many-place decimal) by means of a fraction having the smallest possible denominator. The following convergents of the continued fraction for π, converted into decimals, enable one to see how close the approximations are to the value $\pi = 3{,}14159265 \ldots$:

$$\frac{p_0}{q_0} = 3, \qquad \frac{p_1}{q_1} = \frac{22}{7} = 3{,}14285 \ldots,$$
$$\frac{p_2}{q_2} = \frac{333}{106} = 3{,}141509 \ldots, \quad \frac{p_3}{q_3} = \frac{355}{113} = 3{,}14159292 \ldots.$$

You will observe, moreover, in this example, that the convergents are alternately less than and greater than π. This is true in general, as is well known, that is *the successive convergents of the continued fraction for ω are alternately less than and greater than ω, and enclose it between ever narrowing limits.*

Let us now enliven these considerations with geometric pictures. Confining our attention to positive numbers, let us *mark all those points* in the positive quadrant of the xy plane (see Fig. 8) *which have integral coordinates*, forming thus a so called *point lattice.* Let us examine this lattice, I am tempted to say this "firmament" of points, with our point of view at the origin. The radius vector from 0 to the point $(x = a, y = b)$ has for its equation

$$\frac{x}{y} = \frac{a}{b},$$

and conversely, there are upon every such ray, $x/y = \lambda$, where $\lambda = a/b$ is rational, infinitely many integral points (ma, mb), where m is an arbitrary whole number. Looking from 0, then, one sees points of the lattice *in all rational directions and only in such directions.* The field of view is everywhere "densely" but not completely and continuously filled with "stars". One might be inclined to compare this view with that of the milky way. With the exception of 0 itself there is *not a single integral point lying upon an irrational ray $x/y = \omega$, where ω is irrational,* which is very remarkable. If we recall Dedekind's definition of irrational number, it becomes obvious that such a ray makes a *cut in the field*

of integral points by separating the points into two point sets, one lying to the right of the ray and one to the left. If we inquire how these point sets converge toward our ray $x/y=\omega$, we shall find a very simple relation to the continued fraction for ω. By marking each point ($x=p_\nu$, $y=q_\nu$), corresponding to the convergent p_ν/q_ν, we see that the rays to these points approximate to the ray $x/y=\omega$ better and better, alternately from the left and from the right, just as the numbers p_ν/q_ν approximate to the number ω. Moreover, if one makes use of the known number-theoretic properties of p_ν, q_ν, one finds the following theorem: *Imagine pegs or needles affixed at all the integral points, and wrap a tightly drawn string about the sets of pegs to the right and to the left of the ω-ray, then the vertices of the two convex string-polygons which bound our two point sets will be precisely the points (p_ν, q_ν) whose coordinates are the numerators and denominators of the successive convergents to ω, the left polygon having the even convergents, the right one the odd.* This gives a new, and, one may well say, an extremely graphic definition of a continued fraction. The representation in Fig. 8 corresponds to the example

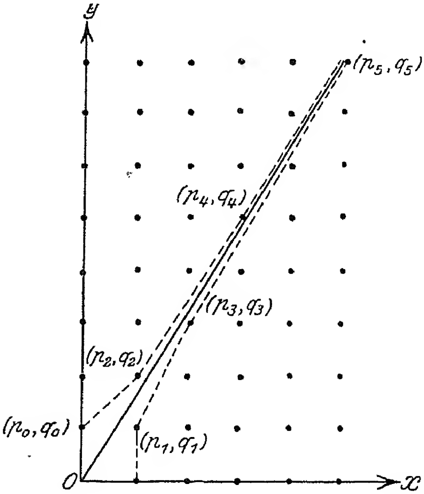


Fig. 8.

$$\omega=\frac{\sqrt{5}-1}{2}=\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{1+\ddots}}}$$

which is the irrationality associated with the regular decagon. In this example, the first few vertices of the two polygons are

left: $p_0=0,\ q_0=1;\quad p_2=1,\ q_2=2;\quad p_4=3,\ q_4=5;\ \ldots$

right: $p_1=1,\ q_1=1;\quad p_3=2,\ q_3=3;\quad p_5=5,\ q_5=8;\ \ldots$

The values p_ν, q_ν for π grow much more rapidly, so that one could hardly draw the corresponding representation. The proof of our theorem, which I cannot give here, can be found in detail on page 43 of in my lithographed lectures.

I shall now pass on to the treatment of the *seventh point*, the *Pythagorean numbers*, where we shall use space perception in a somewhat different form. Instead of the equation:

$$a^2+b^2=c^2, \tag{1}$$

whose integral solutions are sought, let us set:

$$a/c = \xi\,, \qquad b/c = \eta \tag{2}$$

and consider the equation:

$$\xi^2 + \eta^2 = 1\,, \tag{3}$$

with the problem of finding all the *rational number-pairs* ξ, η which satisfy it. Accordingly, we start from the representation of *all rational points* ξ, η (i.e. all points with rational coordinates ξ, η), which will fill the $\xi\,\eta$-plane "densely". $\xi^2 + \eta^2 = 1$ is the equation of the *unit circle* about the origin in this plane. It is our task to see how this *circle threads its way through the dense set of rational points, in particular, to see which of these points it contains.* We know a few such points of old, such as the intercepts with the axes, one of which, $S\,(\xi = -1,\ \eta = 0)$, we shall consider (see Fig. 9). All rays through S are given by the equation

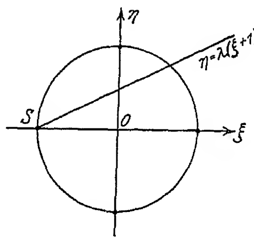


Fig. 9.

$$\eta = \lambda\,(\xi + 1)\,; \tag{4}$$

we call such a ray *rational* or *irrational* according as the parameter λ is rational or not. We have now the double theorem *that every rational point of the circle is projected from S by a rational ray and that every rational ray* (4) *meets the circle in a rational point.* The first half of the theorem is obvious. We prove the second half by substituting from (4) in (3). This gives for the abscissas of the points of intersection the equation

$$\xi^2 + \lambda^2(\xi + 1)^2 = 1$$

or

$$(1 + \lambda^2)\,\xi^2 + 2\lambda^2\xi + \lambda^2 - 1 = 0\,.$$

We know one solution of this equation, $\xi = -1$, which corresponds to the intersection S; for the other, one gets by easy calculation

$$\xi = \frac{1 - \lambda^2}{1 + \lambda^2}\,, \tag{5a}$$

and from (4) the corresponding ordinate

$$\eta = \frac{2\lambda}{1 + \lambda^2}\,. \tag{5b}$$

From (5a) and (5b) it follows that the second intersection is a rational point if λ is rational.

Our double theorem, now fully proved, can be stated also as follows. *All the rational points of the circle are represented by formulas* (5) *if λ is an arbitrary rational number.* This solves our problem and we need only to transform to whole numbers. For this purpose we put

$$\lambda = n/m\,,$$

where n, m are integers and obtain from (5):

$$\xi = \frac{m^2 - n^2}{m^2 + n^2}, \qquad \eta = \frac{2mn}{m^2 + n^2},$$

as the totality of rational solutions of (3). All integral solutions of the original equation (1), i.e., *all Pythagorean numbers are therefore given by the equations*

$$a = m^2 - n^2, \qquad b = 2\,mn, \qquad c = m^2 + n^2;$$

and one obtains the totality of solutions which have no common divisor if m and n take all pairs of relatively prime integral values. We have thus a graphic deduction of a result which usually appears very abstract.

In this connection I should like to discuss the *great Fermat theorem.* It is quite after the manner of the geometers of antiquity that one should generalize the question regarding Pythagorean numbers, from the plane to space of three and more dimensions in the following manner. Is it possible that the sum of the cubes of two integers should be a cube? Or that the sum of two fourth powers should be a fourth power, etc.? In general, *has the equation*

$$x^n + y^n = z^n,$$

where n is an arbitrary integer, solutions which are whole numbers? To this question Fermat gave the answer *no*, in the theorem named after him: *The equation $x^n + y^n = z^n$ has no integral solutions for integral values of n except when $n = 1$ and $n = 2$.* Let me begin with a few historical notes. Fermat lived from 1601 to 1665 and was a parliamentary councillor, i.e., a jurist, in Toulouse. He devoted himself, however, extensively and most fruitfully to mathematics so that he may be counted as one of the greatest of mathematicians. Fermat's name deserves a prominent place among those of the founders of analytic geometry, of infinitesimal calculus, and of the theory of probability. Of special significance however, are his *attainments in the theory of numbers.* All of his results in this field appear as marginal notes on his *copy of Diophantus*, the famous ancient master of number-theory who lived in Alexandria probably about 300 A. D., i. e., about 600 years after Euclid. In this form they were published by his son five years after Fermat's death. Fermat himself had published nothing, but he had, by means of voluminous correspondance with the most significant of his contemporaries, made his discoveries known, although only in part. It was in that edition of Diphantus that the famous theorem with which we are now concerned was found. Fermat wrote concerning it that "he had found a really wonderful proof, but the margin was too narrow to accommodate it"[1]. To this day, no one has succeeded in finding a proof of this theorem!

[1] See the edition issued by the Paris Academy: *Œuvres de Fermat*, vol. I, p. 291. Paris 1891, and vol. III, p. 241. Paris 1896.

In order to orient ourselves somewhat as to its purport, let us inquire, as in the case of $n = 2$, in the first place about the *rational* solutions of the equation:

$$\xi^n + \eta^n = 1,$$

i. e., about the relation of the curve which represents this equation to the totality of the rational points in the $\xi\,\eta$-plane. For $n = 3$ and $n = 4$ the curves have approximately the appearance indicated in Fig. 10, 11 They contain, at least, the points $\xi = 0$, $\eta = 1$ and $\xi = 1$, $\eta = 0$ when

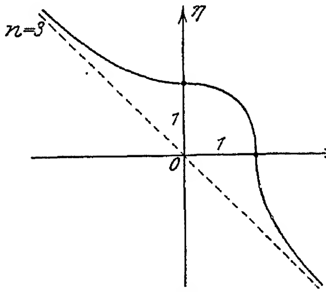


Fig. 10.

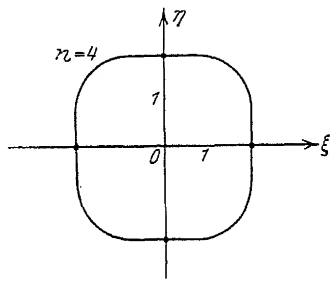


Fig. 11.

$n = 3$, and the points $\xi = 0$, $\eta = \pm 1$ and $\xi = \pm 1$, $\eta = 0$ when $n = 4$. The assertion of Fermat means, now, that these curves, unlike the circle considered above, thread through the dense set of the rational points without passing through a single one, except those just noted.

The interest in this theorem rests on the fact that *all efforts to find a complete proof of it have been, thus far, in vain.* Among those who have attempted proof, one should, above all, mention *Kummer*, who advanced the problem materially by *bringing it into relation with the theory of algebraic numbers*, in particular with the theory of the *n-th roots of unity* (cyclotomic numbers). By using the *n-th* root of 1, $\varepsilon = e^{\frac{2i\pi}{n}}$, we can, indeed, separate $z^n - y^n$ into n linear factors, and we may write the Fermat equation in the form

$$x^n = (z - y)(z - \varepsilon y)(z - \varepsilon^2 y) \ldots (z - \varepsilon^{n-1} y).$$

The problem is therefore reduced to the separation of the *n-th* power of the integer x into n linear factors which shall be built up from two integers z and y and the number ε, in the manner indicated. Kummer developed, for such numbers, theories quite similar to those which have long been known for the case of ordinary integers, theories, that is, which depend on the notions of divisibility and factorization. One speaks, accordingly, of *integral algebraic numbers*, and here, in particular, of cyclotomic numbers, because of the relation of the number ε to the division of the circle. *Fermat's theorem is, then, for Kummer, a theorem on factorization in the domain of algebraic cyclotomic numbers.* From this

theory he tried to deduce a proof of the theorem. He succeeded, in fact, for a very large number of values of n, for example for all values of n below 100. Among the larger numbers, however, there appeared *exceptional values* for which no proof has been found, either by him or by the later mathematicians who continued his investigations.

I must content myself with these remarks. You will find particulars concerning the state of the problem, and concerning Kummer's publications in the Encyclopedia, Vol. I_2, p. 714, at the end of the report by *Hilbert, Theorie der Algebraischen Zahlkörper*. Hilbert himself is among those who have continued and extended the investigations of Kummer[1].

It can indeed hardly be assumed that Fermat's "wonderful proof" lay in this direction. For it is not very likely that he could have operated with algebraic numbers at a time when one was not even certain about the meaning of the imaginary. At that time, also, the theory of numbers was quite undeveloped. It received at the hands of Fermat himself far-reaching stimulation. On the other hand, one cannot assume that a mathematician of Fermat's rank made an error in his proof, although such errors have occurred with the greatest mathematicians. Thus we must indeed believe that he succeeded in his proof by virtue of an especially fortunate simple idea. But as we have not the slightest indication as to the direction in which one could search for that idea, *we shall probably expect a complete proof of Fermat's theorem only through systematic extension of Kummer's work.*

These questions assumed new signifance when our Göttingen Science Association offered a *prize of 100000 marks for the proof of Fermat's theorem*. This was a foundation of the mathematician *Wolfskehl*, who died in 1906. He had probably been interested all his life in Fermat's theorem, and he bequeathed from his large fortune this sum for the fortunate person who should either establish the truth of the theorem of Fermat, or by means of a single example, exhibit its untruth[2]. Such a refutation would, be no simple matter, of course, because the theorem is already proved for exponents below 100 and one would have to start one's calculations with very large numbers.

It will be clear, from my foregoing remarks, how difficult the winning of this prize must seem to the mathematician, who understands the situation and who knows what efforts have been made by Kummer and his successors to prove the theorem. But the *great public* thinks

[[1] A summarized account of the elementary investigations about Fermat's theorem is given in P. Bachmann, *Das Fermatsche Problem*. Berlin 1919.]

[2] The detailed conditions governing competition for this prize (long since become valueless) were published in the Nachrichten d. Ges. d. Wissenschaften zu Göttingen, business announcements 1908, p. 103 et seq., and copied into many other mathematical journals (Sec. e. g. Math. Ann. vol. 66, p. 143; Journal für Mathematik, vol. 134, p. 313).

otherwise. Since the summer of 1907, when the news of the prize was published in the papers (without authorization, by the way) we have received a prodigious heap of alleged "proofs". People of all walks of life, engineers, schoolteachers, clergymen, one banker, many women, have shared in these contributions. The common thing about them all is that they have *no idea of the serious mathematical nature, of the problem.* Moreover, they have made no attempt to inform themselves regarding it, but have trusted to finding the solution by a sudden flash of thought, with the inevitable result that their work is nonsense. One can see what absurdities are brought forth if one reads the numerous critical discussions of such proofs by A. Fleck (who is a practising physician by profession), Ph. Maennchen, and O. Perron, in *Archiv für Mathematik und Physik*[1]. It is amusing to read these wholesale slaughterings, sad as it is that they are necessary. I should like to mention one example, which is related to our treatment of the case $x^2 + y^2 = z^2$. The author seeks a rational parameter representation for the function $x^n + y^n = z^n$ $(n > 2)$, and finds the result, long known from the theory of algebraic functions, that this, unlike the case $n = 2$, is not possible. Now this person overlooks the fact that a non-rational function can very well take on rational values for single rational values of the argument, and he therefore believes that he has proved the Fermat theorem.

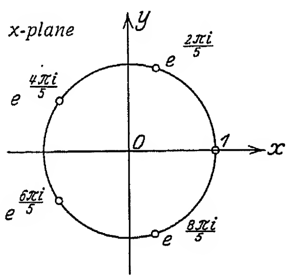


Fig. 12.

With this I close my remarks about Fermat's theorem and come to the *eighth point* of my list, the *problem of the division of the circle.* I shall make use here of operations with complex numbers, $x + iy$, assuming that they are familiar to you, although we shall consider them systematically later on. The *problem is to divide the circle into n equal parts*, or *to construct a regular polygon of n sides.* We identify the circle with the unit circle about the origin of the complex xy-plane and take $x + iy = 1$ as the first of the n points of division (see Fig. 12), in which n is chosen equal to five); then the n complex numbers belonging to the n vertices:

$$z = x + iy = \cos\frac{2k\pi}{n} + i\sin\frac{2k\pi}{n} = e^{\frac{2k\pi i}{n}} \quad (k = 0, 1, \ldots, n-1)$$

satisfy, according to De Moivre's theorem, the equation:

$$z^n = 1,$$

and with this *the problem of the division of the circle is resolved into the solving of this simple algebraic equation.* Since it has the rational root

[1 Vols. XIV, XV, XVI, XVII, XVIII (1901—1911).]

$z = 1$, $z^n - 1$ is divisible by $z - 1$, and there remains for the $n - 1$ other roots the so called *cyclotomic equation*

$$z^{n-1} + z^{n-2} + \cdots + z^2 + z + 1 = 0,$$

an equation of degree $n - 1$, all of whose coefficients are $+1$.

Since ancient times, interest has centered in the question as to *what regular polygons can be constructed with ruler and compasses.* It was known to the ancients that this construction was possible for the numbers $n = 2^h, 3, 5$ (h an arbitrary integer), and likewise for the composite values $n = 2^h \cdot 3 \cdot 5$. Here the problem rested until the end of the eighteenth century when the young Gauss undertook its solution. He found the *desired construction was possible with ruler and compasses for all prime numbers of the form* $p = 2^{(2^\mu)} + 1$, but *for no others.* For the first values $\mu = 0, 1, 2, 3, 4$ this formula yields, in fact, prime numbers, namely

$$3, 5, 17, 257, 65537,$$

of which the first two cases were already known, while the others were new. Of these the *regular polygon of seventeen sides* is especially famous. The fact that it can be constructed with ruler and compasses was first established by Gauss. Moreover, it is not known for what values of μ the above formula yields prime numbers. It has been known, for example, since Euler's time, that for $\mu = 5$ the number is composite. I shall not go farther into details, but rather outline the general conditions, and the significance of this discovery. You will find in Weber-Wellstein details concerning the regular polygon of seventeen sides.

I should like to call to your attention especially the reprint of *Gauss' diary* in the fifty-seventh volume of the *Mathematische Annalen* (1903) and in Volume X, 1 (1917) of Gauss' Works. It is a small, insignificant looking book, which Gauss kept from 1796 on, beginning shortly before his nineteenth birthday. It was precisely the first entry which had to do with the possibility of constructing the polygon of seventeen sides (March 30, 1796); and it was this early important discovery which led Gauss to decide to devote himself to mathematics. The perusal of this diary is of the highest interest for every mathematician, since it permits one, farther on, to follow closely the genesis of Gauss' fundamental discoveries in the field of number theory, of elliptic functions, etc.

The publication of that first great discovery of Gauss appeared as a short communication in the "Jenaer Literaturzeitung" of June 1, 1796, instigated by Gauss' teacher and patron, Hofrat Zimmermann, of Braunschweig, and accompanied by a short personal note by the latter[1]. Gauss published the proof later in his fundamental number-theoretic work,

[1] Also reprinted in Mathematische Annalen, vol. 57, p. 6 (1903); and in Gauss' Works, vol. 10, p. 1 (1917).

Disquisitiones Arithmeticae[1] in 1801; here one finds for the first time the negative part of the theorem, which was lacking in his communication, *that the construction with ruler and compasses is not possible for prime numbers other than those of the form* $2^{2^{\mu}}+1$, e.g., *for* $p = 7$. I shall put before you here an example of this important *proof of impossibility*—the more willingly because there is such a lack of *understanding for proofs of this sort* by the great public. By means of such proofs of impossibility modern mathematics has settled an entire series of famous problems, concerning the solution of which many mathematicians had striven in vain since ancient times. I shall mention, besides the construction of the polygon of seven sides, only the *trisection of an angle* and the *quadrature of the circle* with ruler and compasses. Nevertheless there are surprisingly many persons who devote themselves to these problems without having a glimmering of higher mathematics and without even knowing or understanding the nature of the proof of impossibility. According to their knowledge, which is mostly limited to elementary geometry, they make trials, by drawing, as a rule, auxiliary lines and circles, and multiply these finally in such number that no human being, without undue expenditure of time, can find his way out of the maze and show the author the error in his construction. A reference to the arithmetic proof of impossibility avails little with such persons, since they are amenable, at best, only to a direct consideration of their own "proof" and a direct demonstration of its falsity. Every year brings to every even moderately known mathematician a heap of such consignments, and you also, when you are at your posts, will get such proofs. It is well for you to be prepared in advance for such experiences and to know how to hold your ground. Perhaps it will be well for you, then, if you are master of a definite proof of impossibility in its simplest form.

Accordingly, I should like to give you, in detail, the proof that it is *impossible to construct the heptagon with ruler and compasses in the sense of geometry of precision.* It is well known that every construction with ruler and compasses finds its arithmetic equivalent in a succession of square roots, placed one above another, and, conversely, that one can represent geometrically every such square root by the intersection of lines and circles. This you can easily verify for yourselves. We can formulate our assertion analytically, then, by saying that *the equation of degree six*

$$z^6 + z^5 + z^4 + z^3 + z^2 + z + 1 = 0,$$

which characterizes the regular heptagon, cannot be solved by a succession of square roots in finite number. Now this is a so-called *reciprocal equation,*

[1] Reprinted Works, vol. I.

i. e., it has, for every root z, also $1/z$ as a root. This becomes obvious if we write it in the form:

$$z^3 + z^2 + z + 1 + \frac{1}{z} + \frac{1}{z^2} + \frac{1}{z^3} = 0. \tag{1}$$

We can reduce by half the degree of such an equation, if we take

$$z + \frac{1}{z} = x$$

as a new unknown. By easy calculation, we obtain for x the cubic equation

$$x^3 + x^2 - 2x - 1 = 0, \tag{2}$$

and one sees at once that the equations (1) and (2) are, or are not, both solvable by square roots. Moreover, we can represent x geometrically in connection with the construction of the heptagon. For, if we consider the unit circle in the complex plane, we see easily that the following relations are obvious. If one designates by $\varphi = \frac{2\pi}{7}$ the central angle of the regular heptagon, and remembers that $z = \cos\varphi + i \sin\varphi$ and $\frac{1}{z} = \cos\varphi - i\sin\varphi$ are the two vertices of the heptagon nearest to $x = 1$, then $x = z + \frac{1}{z} = 2\cos\varphi$ (Fig. 13). Thus, if one knows x, one can at once construct the heptagon.

We must now show *that the cubic equation* (2) *cannot be solved by square roots*. The proof *falls into an arithmetic and an algebraic part.*

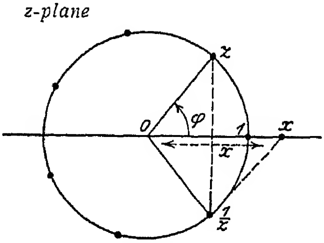


Fig. 13.

We shall start by showing that the equation (2) is *irreducible*, i. e. that its left side cannot be separated into two factors whose coefficients are rational numbers. Let us assume that the equation is reducible. Then its left side must have a linear factor with rational coefficients, and hence it must vanish for a rational number p/q, where p and q are integers without a common divisor. But that means that $p^3 + p^2 q - 2pq^2 - q^3 = 0$, or that p^3, and therefore p itself, is divisible by q. In the same way it follows that q^3, and hence q, must be divisible by p. Consequently $p = \pm q$ and the equation (2) must have the root $x = \pm 1$. But inspection shows that this is not the case.

The *second part of the proof* consists, in showing *that an irreducible cubic equation with rational coefficients is not solvable by square roots*. It is essentially *algebraic* in nature, but because of the connection I shall give it here. Let us make the assertion in positive form. *If a cubic equation with rational coefficients* A, B, C:

$$f(x) = x^3 + Ax^2 + Bx + C = 0 \tag{8}$$

can be solved by square roots, it must have a rational root, i. e., it is reducible. For the existence of a rational root α is equivalent to the existence of a *rational* factor $x - \alpha$ of $f(x)$ and thus to reducibility. It is most important that this proof be preceded by a *classification of all expressions that can be built up with square roots,* or, more precisely, of *all expressions that can be built up with square roots and rational numbers, in finite number, by means of rational operations.* A concrete example of such a number is

$$\alpha = \frac{\sqrt{a + \sqrt{b} + \sqrt{c}}}{\sqrt{d + \sqrt{e + \sqrt{f}}}},$$

where $a, b, \ldots, f$ are rational numbers. Of course we are talking only about square roots which cannot be extracted rationally. All others must be simplified. Every such expression is a rational function of a certain number of square roots. In our example there are three. We shall first consider *a single such square root,* whose radicand, however, may have a form as complicated as one pleases. *By its "order" we shall understand the largest number of root signs which appear in it, one above another.* In the preceding example, α, the roots of the numerator have the orders 2 and 1, respectively, while that of the denominator has the order 3.

In the case of a *general square root expression* we examine the *orders of the different "simple square root expressions" of the sort just discussed,* out of which the general expression is rationally constructed, and we *designate the largest among them as the order μ of the expression in question.* In our example, $\mu = 3$. Now several "simple square root expressions" of order μ might appear in our expression and we consider *their number, n, the "number of terms"* of order μ, as a second characteristic. This number is thought of as so determined that *no one of the n simple expressions of order μ can be rationally expressed in terms of the others of order μ, or of lower order.* For example, the expression of order 1

$$\sqrt{2} + \sqrt{3} + \sqrt{6}$$

has 2, not 3, as the "number of terms" since $\sqrt{6} = \sqrt{2} \cdot \sqrt{3}$. The example α given above has $n = 1$.

We have thus assigned to every square root expression two finite numbers μ, n which we combine in the *symbol (μ, n) as the "characteristic" or "rank" of the root expression. When two root expressions have different orders we assign a lower rank to the one of lower order; when the orders are the same, the lower number of terms determines the lower rank.*

Now let us suppose that a *root* x_1 of the cubic equation (8) is expressible by means of square roots; and, to be explicit, by means of an *expression of rank* (μ, n). Selecting one of the n terms $\sqrt{R}$ of rank μ, let x be written in the form

$$x_1 = \frac{\alpha + \beta\sqrt{R}}{\gamma + \delta\sqrt{R}},$$

where α, β, γ, δ *contain at most* $n-1$ *terms of order* μ and where R *is of order* $\mu - 1$. Here $\gamma - \delta\sqrt{R}$ is certainly different from zero; for $\gamma - \delta\sqrt{R} = 0$ would imply either $\delta = \gamma = 0$, which is obviously impossible, or $\sqrt{R} = \gamma : \delta$, i.e., $\sqrt{R}$ would be rationally expressible by means of the other $(n-1)$ terms of order μ, which appear in x, and hence it would be superfluous. Multiplying numerator and denominator by $\gamma - \delta\sqrt{R}$, we find

$$x_1 = \frac{(\alpha + \beta\sqrt{R})(\gamma - \delta\sqrt{R})}{\gamma^2 - \delta^2 \cdot R} = P + Q\sqrt{R},$$

where P, Q are *rational functions of* $\alpha, \beta, \gamma, \delta$, that is, they contain at most $(n-1)$ terms of order μ, and, besides, only those of order $\mu - 1$, so that they have *at most the rank* $(\mu, n-1)$. Substituting this value of x in (8), we get

$$f(x_1) = \left(P + Q\sqrt{R}\right)^3 + A\left(P + Q\sqrt{R}\right)^2 + B\left(P + Q\sqrt{R}\right) + C = 0,$$

and when we remove parentheses we obtain a relation of the form

$$f(x_1) = M + N\sqrt{R} = 0,$$

where M, N are polynomials in P, Q, R, that is, rational functions of α, β, γ, δ, R. If $N \neq 0$, we should have $\sqrt{R} = -M/N$, i.e., $\sqrt{R}$ would be expressible rationally in terms of α, β, γ, δ, R, that is, by means of the other $(n-1)$ terms of order μ and others of lower order. But that is impossible, as remarked above, according to the hypothesis. Thus it follows necessarily that $N = 0$ and hence also $M = 0$. From this we may conclude, that

$$x_2 = P - Q\sqrt{R}$$

is also a root of the cubic equation (8). For a comparison with the last equations yields at once

$$f(x_2) = M - N\sqrt{R} = 0.$$

The proof may now be finished very simply and surprisingly. If x_3 is *the third root* of our cubic equation, we have

$$x_1 + x_2 + x_3 = -A,$$

and hence
$$x_3 = -A - (x_1 + x_2) = -A - 2P$$

is of the same rank as P and *therefore certainly of lower rank than* x_1.

If x_3 is itself rational, our theorem is proved. If not, we can make it the starting point of the same series of deductions. It appears that, in the case of the other roots, the higher rank must have been an illusion, so *that, in particular, one of them has, actually, lower rank than* x_3. If we keep this up, back and forth among the roots, we see, each time, that the rank is really lower than we had thought. We must, then, of necessity, come finally to a root with the order $\mu = 0$. *This demon-*

strates the existence of a rational root of the cubic equation. We cannot continue our procedure beyond this point. The two other roots must then be, either themselves rational, or else of the form $P = Q\sqrt{R}$, where P, Q, R are rational numbers. Hence we have shown *that $f(x)$ separates into a quadratic and a linear rational factor and is therefore reducible. Every irreducible cubic equation, and in particular, our equation for the regular heptagon, is insoluble by means of square roots.* The proof is therefore complete *that the regular heptagon cannot be constructed with ruler and compasses.*

You observe how simply and obviously this proof proceeds, and how little knowledge it really presupposes. For all that, some of the steps, especially the explanation of the classification of square root expressions, demand a certain measure of mathematical abstraction. Whether the proof is simple enough to convince one of those mathematical laymen, mentioned above, of the futility of his attemps at an elementary geometric proof, I do not presume to decide. Nevertheless one should try to explain the proof slowly and clearly to such a person.

In conclusion, I shall mention some of the *literature on the question of regular polygons* together with some, on the broader question of geometric constructibility in general which we have touched upon on this occasion. First of all, there is again *Weber-Wellstein* I (Sections 17 and 18 in the fourth edition). Next let me mention the souvenir booklet *Vorträge über ausgewählte Fragen der Elementargeometrie*[1*] which I prepared in 1895, on the occasion of a gathering of teachers in Göttingen. I might mention, as a more detailed and comprehensive substitute for this little book (which is out of print) the German translation, *Fragen der Elementargeometrie*[2**], of a compilation by *F. Enriques* in Bologna, where you will find information on all allied questions.

I leave now the discussion of number theory, reserving the last point, the transcendence of π, for the conclusion of this course of lectures, and turn, in the next chapter, to our final extension of the number system.

IV. Complex Numbers.

1. Ordinary Complex Numbers

Let me give, as a preliminary, *some historical facts.* Imaginary numbers are said to have been used first, incidentally, to be sure, by *Cardan* in 1545, in his solution of the cubic equation. As for the further

[1] Worked up by F. Tägert. Leipzig 1895.

[2] Teil II: *Die geometrischen Aufgaben, ihre Lösung und Lösbarkeit.* Deutsch von H. Fleischer. Leipzig 1907. [2. Aufl. 1923.]—See also *Young, J. W. A., Monographs on Topics in Modern Mathematics.*

* Translation by Beman and Smith: *Famous Problem of Geometry.* Ginn, reprinted by Stechert, New York.

** *Problems of Elementary Geometry.*

development, we can make the same statement as in the case of negative numbers, *that imaginary numbers made their own way into arithmetic calculation without the approval, and even against the desires of individual mathematicians, and obtained wider circulation only gradually and to the extent to which they showed themselves useful.* Meanwhile the mathematicians were not altogether happy about it. Imaginary numbers long retained a somewhat *mystic* coloring, just as they have today for every pupil who hears for the first time about that remarkable $i = \sqrt{-1}$. As evidence, I mention a very significant utterance by *Leibniz* in the year 1702, "Imaginary numbers are a fine and wonderful refuge of the divine spirit, almost an amphibian between being and non-being". In the eighteenth century, the notion involved was indeed by no means cleared up, although *Euler*, above all, *recognized their fundamental significance for the theory of functions.* In 1748 Euler set up that remarkable relation:

$$e^{ix} = \cos x + i \sin x$$

by means of which one recognizes the fundamental relationship among the kinds of functions which appear in elementary analysis. The *nineteenth century finally brought the clear understanding of the nature of complex numbers.* In the first place, we must emphasize here the *geometric interpretation* to which various investigators were led about the end of the century. It will suffice if I mention the man who certainly went deepest into the essence of the thing and who exercised the most lasting influence upon the public, namely Gauss. As his diary, mentioned above, proves incontrovertibly, he was, in 1797, already in full possession of that interpretation, although, to be sure, it was published very much later. The second achievement of the nineteenth century is the creation of a *purely formal foundation* for complex numbers, which reduces them to dependence upon real numbers. This originated with English mathematicians of the thirties, the details of which I shall omit here, but which you will find in Hankel's book, mentioned above.

Let me now explain these *two prevailing foundation methods.* We shall take first the *purely formal standpoint,* from which the consistency of the rules of operation among themselves, rather than the meaning of the objects, guarantees the correctness of the concepts. According to this view, complex numbers are introduced in the following manner, which precludes every trace of the mysterious.

1. The complex number $x + iy$ is the *combination of two real numbers* x, y, that is, a *number-pair,* concerning which one adopts the conventions which follow.

2. Two complex numbers $x + iy, x' + iy'$ are called *equal* when $x = x', y = y'$.

3. Addition and subtraction are defined by the relation

$$(x + iy) \pm (x' + iy') = (x \pm x') + i(y \pm y').$$

All the *rules of addition* follow from this, as is easily verified. The *monotonic law* alone loses its validity in its original form, since complex numbers, by their nature, do not have the same simple order in which natural or real numbers appear by virtue of their magnitude. For the sake of brevity I shall not discuss the modified form which this gives to the monotonic law.

4. We stipulate that in *multiplication* one operates as with ordinary letters, except that one always puts $i^2 = -1$; in particular, that

$$(x + iy)(x' + iy') = (xx' - yy') + i(xy' + x'y).$$

It is easy to see that, with this, *all the laws of multiplication hold, with the exception of the monotonic law, which does not enter into consideration.*

5. *Division* is defined as the *inverse of multiplication*; in particular, we may easily verify that

$$\frac{1}{x + iy} = \frac{x}{x^2 + y^2} - i\frac{y}{x^2 + y^2}.$$

This number always exists except for $x = y = 0$, i.e., *division by zero has the same exceptional place here as in the domain of real numbers.*

It follows from this that operations with complex numbers cannot lead to contradictions, since they depend exclusively upon real numbers and known operations with them. We shall assume here that these are devoid of contradiction.

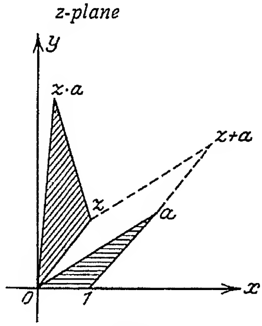


Fig. 14.

Besides this purely formal treatment, we should of course like to have a geometric, or otherwise visual, interpretation of complex numbers and of operations with them, in which we might see a *graphical foundation of consistency*. This is supplied by common geometric interpretation, which, as you all know and as we have already mentioned, *looks upon the totality of points* (x, y) *of the plane in an* xy*-coordinate system as representing the totality of complex numbers* $z = x + iy$. The sum of two numbers z, a follows by means of the familiar *parallelogram construction* with the two corresponding points and the origin 0, while the product $z \cdot a$ is obtained by constructing on the segment $0z$ a triangle similar to $a01$, where 1 is the point $(x = 1, y = 0)$ (Fig. 14). In brief, *addition* $z' = z + a$ *is represented by a translation of the plane into itself, multiplication* $z' = z \cdot a$ *by a similarity transformation,* i.e., *by a turning and a stretching, the origin remaining fixed.* From the order of the points

in the plane, considered as representatives of complex numbers, one sees at once what takes the place here of the monotonic laws for real numbers. These suggestions will suffice, I hope, to recall the subject clearly to your memory.

I must call to your attention the place in Gauss in which this foundation of complex numbers, by means of their geometric interpretation, is set out with full emphasis, since it was this which first exhibited the general importance of complex numbers. In the year 1831 Gauss' researches carried him into the theory especially of *integral complex numbers* $a + ib$, where a, b are real integers, in which he developed for the new numbers the theorems of ordinary *number theory* concerning prime factors, quadratic and biquadratic residues, etc. We mentioned such generalizations of number theory, in connection with our discussion of Fermat's theorem. In his own abstract[1] of this paper Gauss expresses himself concerning what he calls the "true metaphysics of imaginary numbers". For him, the right to operate with complex numbers is justified by the geometric interpretation which one gives to them and to the operations with them. Thus he takes *by no means the formal standpoint.* Moreover, these long, beautifully written expositions of Gauss are extremely well worth reading. I mention here, also, that Gauss proposes the clearer word "complex", instead of "imaginary", a name that has, in fact, been adopted.

2. Higher Complex Numbers, especially Quaternions

It has occurred to everyone who has worked seriously with complex numbers to ask if we cannot set up other, higher, complex numbers, with more ne wunits than the one i and if we cannot operate with them logically. Positive results in this direction were obtained about 1840 by *H. Grassmann,* in Stettin, and *W. R. Hamilton,* in Dublin, independently of each other. We shall examine the invention of Hamilton, the *calculus of quaternions,* somewhat carefully later on. For the present let us look at the general problem.

We can look upon the ordinary complex number $x + iy$ as a *linear combination*

$$x \cdot 1 + y \cdot i$$

formed from two different "*units*" 1 and i, by means of the *real parameters* x and y. Similarly, let us now imagine an arbitrary number, n, of units $e_1, e_2, \ldots, e_n$ all different from one another, and let us call the totatily of combinations of the form $x = x_1e_1 + x_2e_2 + \ldots, + x_ne_n$ a *higher complex number system* formed from them with n arbitrary real numbers $x_1, x_2, \ldots, x_n$. If there are given two such numbers, say x, defined above, and

$$y = y_1e_1 + y_2e_2 + \ldots, + y_ne_n,$$

[1] See Werke, vol. II.

it is nearly obvious that we should call them *equal when, and only when, the coefficients of the individual units, the so called "components" of the number, are equal in pairs*

$$x_1 = y_1, \; x_2 = y_2, \; \ldots \; x_n = y_n.$$

The definition of addition and subtraction, which reduces these operations simply to the *addition and subtraction of the components*,

$$x \pm y = (x_1 \pm y_1)e_1 + (x_2 \pm y_2)e_2 + \ldots, + (x_n \pm y_n)\, e_n,$$

is equally obvious.

The matter is more difficult and more interesting in the case of *multiplication*. To start with, we shall proceed according to the general rule for multiplying letters, i. e., multiply each i-th term of x by every k-th term of y $(i, k = 1, 2, \ldots, n)$. This gives:

$$x \cdot y = \sum_{(i,\, k=1,\, \ldots,\, n)} x_i y_k e_i e_k .$$

In order that this expression should be a number in our system, one must have a rule which represents the *products* $e_i \cdot e_k$ *as complex numbers of the system,* i. e., as linear combinations of the units. Thus one must have n^2 equations of the form:

$$e_i e_k = \sum_{(l=1,\, \ldots,\, n)} c_{ikl} \cdot e_l . \qquad (i,\, k = 1, \ldots, n)$$

Then we may say that the number

$$x \cdot y = \sum_{(l=1,\, \ldots,\, n)} \Big\{ \sum_{(i,\, k=1,\, \ldots,\, n)} x_i y_k c_{ikl} \Big\}\, e_l$$

will always belong to our complex number system. *Each particular complex number system is characterized by the method of determining this rule for multiplication, i.e., by the table of the coefficients* C_{ikl}.

If one now defines *division as the operation inverse to multiplication*, it turns out that, under this general arrangement, division *is not always uniquely possible*, even when the divisor does not vanish. For, the determination of y from $x \cdot y = z$ requires the solution of the n linear equations $\sum\limits_{i,\, k} x_i y_k C_{ikl} = z_l$ for the n unknowns $y_1, \ldots, y_n$, and these would have either no solution, or infinitely many solutions, if their determinant happened to vanish. Moreover, all the z_l may be zero even when not all the x_i or not all the y_k vanish, i. e., *the product of two numbers can vanish without either factor being zero.* It is only by a skillful special choice of the numbers C_{ikl} that one can bring about accord here with the behavior of ordinary numbers. To be sure, a closer investigation *shows, when $n > 2$, that, to attain this, we must sacrifice one of the other rules of operation.* We choose as the rule that fails to be satisfied, one which appears less important under the circumstances.

Let us now follow up these general explanations by a more detailed discussion of *quaternions* as the example which, by reason of its applications in physics and mathematics, constitutes the *most important higher complex number system*. As the name indicates, these are *four-term numbers* ($n = 4$); as a sub-class, they include the *three-term vectors*, which are generally known today, and which are sometimes discussed in the schools.

As the first of the four units with which we shall construct quaternions, we shall select the *real unit* 1, (as in the case of ordinary complex numbers). We ordinarily denote the other three units, as did Hamilton, by i, j, k, so that the general from of the quaternion is

$$p = d + ia + jb + kc,$$

where a, b, c, d *are real parameters, the coefficients of the quaternion*. We call the first component, the one which is multiplied by 1, and which corresponds to the real part of the common complex number, the *"scalar part" of the quaternion*, the aggregate $ai + bj + ck$ of the other three terms its *"vector part"*.

The addition of quaternions follows from the preceding general remarks. I shall give an obvious *geometric interpretation*, which goes back to that interpretation of vectors which is familiar to you. We imagine the *segment*, corresponding to the vector part of p, and having the projections a, b, c on the coordinate axes, as loaded with a *weight* equal to the scalar part. Then addition of p and $p' = d' + ia' + jb' + kc'$ is accomplished by constructing the resultant of the two segments, according to the well known parallelogram law of vector addition (see Fig. 15), and then loading it with the sum of the weights, for this would then in fact represent the quaternion:

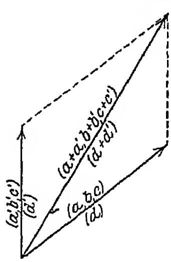


Fig. 15.

$$(1) \quad p + p' = (d + d') + i(a + a') + j(b + b') + k(c + c').$$

We come first to specific properties of quaternions when we turn to *multiplication*. As we saw in the general case, these properties must be implicit in the *conventions adopted as to the products of the units*. To begin with, I shall indicate the quaternions to which *Hamilton* equated the sixteen products of two units each. As its symbol indicates, we shall operate with the first unit 1 as with the real number 1, so that:

$$(2a) \quad 1^2 = 1, \quad i \cdot 1 = 1 \cdot i = i, \quad j \cdot 1 = 1 \cdot j = j, \quad k \cdot 1 = 1 \cdot k = k.$$

As something essentially new, however, we agree that, for the squares of the other units:

$$(2b) \qquad i^2 = j^2 = k^2 = -1,$$

and for their binary products:

(2c) $$jk = +i, \quad ki = j, \quad ij = +k$$

whereas for the inverted position of the factors:

(2d) $$kj = -i, \quad ik = -j, \quad ji = -k.$$

One is struck here by the fact that the *commutative law for multiplication is not obeyed*. This is the inconvenience in quaternions which one must accept in order to rescue the uniqueness of division, as well as the theorem that a product should vanish only when one of the factors vanishes. *We shall show at once that not only this theorem but also all the other laws of addition and multiplication remain valid, with this one exception, in other words, that these simple agreements are very expedient.*

We construct, first, the *product of two general quaternions*

$$p = d + ia + jb + kc \quad \text{and} \quad q = w + ix + jy + kz.$$

Let us start from the equation

$$q' = p \cdot q = (d + ia + jb + kc) \cdot (w + ix + jy + kz);$$

and let us multiply out term by term. In carrying out this multiplication, we must note the order in the case of the units i, j, k. We must follow the commutative law for products composed of the components a, b, c, d, and for products of components and one unit, we must replace the products of units in accordance with our multiplication table, and we must then collect the terms having the same unit. We must then collect the terms having the same unit. We then have

(3) $$\left.\begin{aligned} q' = pq = w' + ix' + jy' + kz' = {} & (dw - ax - by - cz) \\ & + i(aw + dx \underline{+ bz - cy}) \\ & + j(bw + dy \underline{+ cx - az}) \\ & + k(cw + dz \underline{+ ay - bx}). \end{aligned}\right\}$$

The components of the product quaternion are thus definite simple *bilinear combinations* of the components of the two factors. If we invert the order of the factors, the six underscored terms change their signs, so that $q \cdot p$, *in general, is different from* $p \cdot q$, and the difference is more than a change of sign as was the case with the individual units.

Although the commutative law fails for multiplication, the *distributive and associative laws hold without change.* For, if we construct on the one hand $p(q + q_1)$, on the other $pq + pq_1$ by multiplying out formally without replacing the products of the units, we must, of necessity, get identical results, and no change can be brought about by then using the multiplication table. Further, the associative law must hold in general, if it holds for the multiplication of the units.

But this follows at once from the multiplication table, as the following example shows:

$$(ij)\,k = i\,(jk)\,.$$

In fact, we have:

$$(ij)k = k\cdot k = -1\,,$$

and

$$i(jk) = i\cdot i = -1\,.$$

We shall now take up *division*. It will suffice to show *that for every quaternion* $p = d + ia + jb + kc$ *there is a definite second one*, q, *such that*:

$$p\cdot q = 1\,.$$

We shall denote q appropriately by $1/p$. Division in general can be reduced easily to this special case, as we shall show later. In order to determine q, let us put, in equation (3),

$$q' = 1 = 1 + 0\cdot i + 0\cdot j + 0\cdot k,$$

and obtain, by equating components, the following four equations for four unknown components x, y, z, w of q:

$$\begin{aligned} dw - ax - by - cz &= 1 \\ aw + dx - cy + bz &= 0 \\ bw + cx + dy - az &= 0 \\ cw - bx + ay + dz &= 0\,. \end{aligned}$$

The solvability of such a system of equations depends, as is well known, upon its determinant, which, in the case before us, is a skew symmetric determinant, in which all the elements of the principal diagonal are the same, and all the pairs of elements which are symmetrically placed with respect to that diagonal are equal and opposite in sign. According to the theory of determinants, such determinants are easily calculated; and we find

$$\begin{vmatrix} d & -a & -b & -c \\ a & d & -c & b \\ b & c & d & -a \\ c & -b & a & d \end{vmatrix} = (a^2 + b^2 + c^2 + d^2)^2\,.$$

By direct calculation this result can be easily verified. The real *elegance of Hamilton's conventions* depends upon this result, that the determinant is a power of the sum of squares of the four components of p; for it follows *that the determinant is always different from zero except when* $a = b = c = d = 0$. With this one self evident exception ($p = 0$), the equations are *uniquely solvable* and the *reciprocal quaternion* q *is uniquely determined*.

The quantity

$$T = \sqrt{a^2 + b^2 + c^2 + d^2}$$

plays an important role in the theory, and is called the *tensor of* p. It is easy to show that these unique solutions are

$$x = -\frac{a}{T^2}, \qquad y = -\frac{b}{T^2}, \qquad z = -\frac{c}{T^2}, \qquad w = \frac{d}{T^2}$$

so that we have as the final result

$$\frac{1}{p} = \frac{1}{d + ia + jb + kc} = \frac{d - ia - jb - kc}{a^2 + b^2 + c^2 + d^2}.$$

If we introduce the *conjugate value* of p, as in ordinary complex numbers:

$$\bar{p} = d - ia - jb - kc,$$

we can write the last formula in the form

$$\frac{1}{p} = \frac{\bar{p}}{T^2}$$

or

$$p \cdot \bar{p} = T^2 = a^2 + b^2 + c^2 + d^2.$$

These formulas which are immediate generalizations of certain properties of ordinary complex numbers. Since p is also the number conjugate to $\bar{p}$, it follows also that:

$$\bar{p} \cdot p = T^2,$$

so that the commutative law holds in this special case.

The general problem of division can now be solved. For, from the equation

$$p \cdot q = q',$$

it follows, by multiplication by $1/p$, that

$$q = \frac{1}{p} \cdot q' = \frac{\bar{p}}{T^2} \cdot q',$$

whereas the equation

$$q \cdot p = q',$$

which one gets by changing the order of the factors, has the solution

$$q = q' \cdot \frac{1}{p} = q' \cdot \frac{\bar{p}}{T^2}.$$

This solution is different, in general, from the other.

Now we must inquire whether there is a *geometric interpretation of quaternions* in which these operations, together with their laws, appear in a natural form. In order to arrive at it, we start with the special case in which *both factors reduce to simple vectors*, i. e., in which the

scalar parts w, d, are zero. The formula (3) for multiplication then becomes

$$q' = p \cdot q = (ia + jb + kc)(ix + jy + kz)$$
$$= -(ax + by + cz) + i(bz - cy) + j(cx - az) + k(ay - bx),$$

i. e., *when each of two quaternions reduces to a vector, their product consists of a scalar and a vector part.* We can easily bring these two parts into relation with the different kinds of *vector multiplication* which are in use. The notions of vector calculus, which is far more wide spread than quaternion calculus, go back to *Grassmann*, although the word *vector* is of English origin. The two kinds of vector product with which one usually operates are designated now, mostly, by *inner* (*scalar*) *product* $ax + by + cz$ (i. e., the scalar part of the above quaternion product, except for the sign), and *outer* (*vector*) *product* $i(bz - cy) + j(cx - az) + k(ay - bx)$, (i. e., the vector part of the quaternion product. We shall give a geometric interpretation of each part separately:

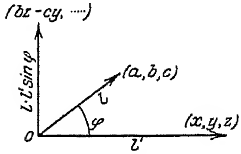


Fig. 16.

Let us lay off both vectors (a, b, c) and (x, y, z), as segments, from the origin O (Fig. 16). They terminate in the points (a, b, c) and (x, y, z) respectively, and have the lengths $l = \sqrt{a^2 + b^2 + c^2}$ and $l' = \sqrt{x^2 + y^2 + z^2}$. If φ is the angle between these two segments, then, according to well known formulas of analytic geometry, which I do not need to develop here, the *inner product* is:

$$ax + by + cz = l \cdot l' \cdot \cos\varphi;$$

and the *outer product*, on the other hand, is itself a *vector*, which, as is easily seen, is *perpendicular to the plane of l and l'* and has the *length* $l \cdot l' \cdot \sin\varphi$.

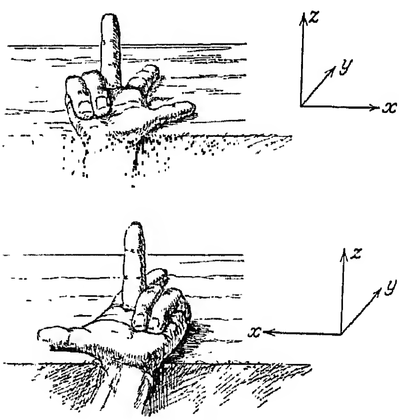


Fig. 17.

It is essential now to decide as to the *sense of the product vector*, i. e., toward which side of the plane determined by l and l' one is to lay off this vector. This sense is *different according to the coordinate system* which one chooses. As you know, one can choose *two rectangular coordinate systems which are not congruent*, i. e., *which cannot be made to coincide with one another*, by holding, say, the y- and the z-axis fixed and reversing the sense of the x-axis. These systems are then *symmetric to each other*, like *the right and the left hand* (Fig. 17). The distinction between them can be borne in mind by the following rule: *In the one*

system, the x, y, and z axis lie like the outstretched thumb, fore finger and middle finger, respectively, of the right hand; in the other, like the same fingers of the left hand. These two systems are used confusedly in the literature; different habits obtain in different countries, in different fields, and, finally, with different writers, or even with the same writer. Let us now examine the simplest case, where $p = i$, $q = j$, these being the unit lengths laid off on the x and y axis. Then, since $i \cdot j = k$, the outer vector product is the unit length laid off on the z-axis. (See Fig. 18.) Now one can transform i and j continuously into two arbitrary vectors p and q so that k transforms continuously into the vector component of $p \cdot q$ without going through zero. Consequently the *first factor, the second factor, and the vector product must always lie, with respect to each other, like the x, y, and z-axis of the system of coordinates, i. e., right-handed* (as in Fig. 18) *or left-handed* (as in Fig. 16), *according to the choice of coordinate system.* (In Germany, now, the choice indicated in Fig. 18 is customary.)

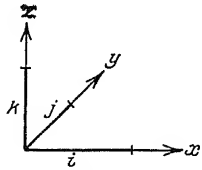


Fig. 18.

I should like to add a few words concerning the much disputed *question of notation in vector analysis.* There are, namely, a great many different symbols used for each of the vector operations, and it has been impossible, thus far, to bring about a generally accepted notation. At the meeting of natural scientists at Kassel (1903) a commission was set up for this purpose. Its members, however, were not able even to come to a complete understanding among themselves. Since their intentions were good, however, each member was willing to meet the others part way, so that the only result was that about three new notations came into existence! My experience in such things inclines me to the belief that real agreement could be brought about only if important material interests stood behind it. It was only after such pressure that, in 1881, the uniform system of measures according to volts, amperes, and ohms was generally adopted in electrotechnics and afterward settled by public legislation, due to the fact that industry was in urgent need of such uniformity as a basis for all of its calculations. But there are no such strong material interests behind vector calculus, as yet, and hence one must agree, for better or worse, to let every mathematician cling to the notation which he finds the most convenient, or—if he is dogmatically inclined—the only correct one.

3. Quaternion Multiplication — Rotation and Expansion

Before we proceed to the consideration of the geometric meaning of multiplication of general quaternions, let us consider the following question. Let us consider the product $q' = p \cdot q$ of two quaternions p and q, and let us replace p and q by their conjugates $\bar{p}$ and $\bar{q}$, that

is, let us change the signs of a, b, c, x, y, z. Then the scalar part of the product, as given in (3), p. 61, remains unchanged, and only those factors of i, j, k which are not underscored will change sign. On the other hand, if we also reverse the order of the factors $\bar{p}$ and $\bar{q}$, the factors of i, j, k which are underscored will change sign. Hence the product $\bar{q}' = \bar{q} \cdot \bar{p}$ is precisely the conjugate of the original product q'; and we have

$$q' = p \cdot q, \qquad \bar{q}' = \bar{q} \cdot \bar{p},$$

where $\bar{q}'$ is the conjugate of q'. If we multiply these two equations together, we obtain

$$q' \cdot \bar{q}' = p \cdot q \cdot \bar{q} \cdot \bar{p}.$$

In this equation the order of the factors is essential, since the commutative law does not hold. We may apply the associative law, however, and we may write

$$q' \cdot \bar{q}' = p \cdot (q \cdot \bar{q}) \cdot \bar{p}.$$

Since we have, by p. 63,

$$q \cdot \bar{q} = x^2 + y^2 + z^2 + w^2,$$

we may write

$$w'^2 + x'^2 + y'^2 + z'^2 = p\,(w^2 + x^2 + y^2 + z^2)\,\bar{p}.$$

The middle factor on the right is a scalar, and the commutative law does hold for multiplication of a scalar by a quaternion, since $M \cdot p = Md + i(Ma) + j(Mb) + k(Mc) = pM$. Hence we have

$$w'^2 + x'^2 + y'^2 + z'^2 = p\bar{p}\,(w^2 + x^2 + y^2 + z^2),$$

and, since $p \cdot \bar{p}$ is the square of the tensor of p, we find[1]

$$\text{(I)} \quad w'^2 + x'^2 + y'^2 + z'^2 = (d^2 + a^2 + b^2 + c^2)\,(w^2 + x^2 + y^2 + z^2),$$

that is, *the tensor of the product of two quaternions is equal to the product of the tensors of the factors*. This formula can be obtained also by direct calculation, by taking the values of w', x', y', z' from the formula for a product given on p. 61.

We shall now represent a quaternion as the segment joining the origin of a four-dimensional space to the point (x, y, z, w), in a manner exactly analogous to the representation of a vector in three-dimensional space. It is no longer necessary to apologize for making use of four-dimensional space, as was the custom when I was a student. All of you are fully aware that no metaphysical meaning is intended, and that higher dimensional space is nothing more than a convenient mathematical expression which permits us to use terminology analogous to that of

[1] This formula, in all that is essential, occurs in Lagrange's works.

actual space representation. If we regard p as a constant, that is, if we regard a, b, c, d as constants, the quaternion equation

$$q' = p \cdot q$$

represents a certain *linear tranformation* of the points (x, y, z, w) of the four-dimensional space into the points (x', y', z', w'), since the equation assigns to every four-dimensional vector q another vector q' linearly. The explicit equations for this transformation, i.e., the expressions for x', y', z', w' as linear functions of x, y, z, w, may be obtained by comparison of the coefficients of the product formula (3), p. 61. The tensor equation (I) shows that the distance of any point from the origin, $\sqrt{x^2 + y^2 + z^2 + w^2}$, is multiplied by the same constant factor $T = \sqrt{a^2 + b^2 + c^2 + d^2}$, for all points of the space. Finally, by p. 62, the determinant of the linear transformation is surely positive.

It is shown in analytic geometry of three-dimensional space that if a linear transformation of the coordinates x, y, z is *orthogonal* (that is, if it carries the expression $x^2 + y^2 + z^2$ into itself), and if the determinant of the transformation is positive, the transformation represents a *rotation about the origin*. Conversely, *any* rotation can be obtained in this manner. If the linear transformation carries $x^2 + y^2 + z^2$ into the similar expression in x', y', z' multiplied by a constant factor T^2, however, and if the determinant is positive, the transformation represents a *rotation about the origin combined with an expansion in the ratio T about the origin*, or, briefly, *a rotation and expansion.*

The facts just mentioned for three-dimensional space may be extended to four-dimensional space. We shall say that our transformation of four-dimensional space represents in precisely the same sense a *rotation and expansion about the origin.* It is easy to see, however, that in this case we do *not* obtain the most general rotation and expansion about the origin. For our transformation contains only four arbitrary constants, namely, the components a, b, c, d of p, whereas, as we shall show immediately, the most general rotation and expansion about the origin in the four-dimensional space R_4 contains seven arbitrary constants. Indeed, in order that the general linear transformation should be a rotation and expansion, we must have

$$x'^2 + y'^2 + z'^2 + w'^2 = T^2(x^2 + y^2 + z^2 + w^2).$$

If we replace x', y', z', w' by linear integral functions of x, y, z, w, we obtain a quadratic form in four variables, which contains $(4 \cdot 5)/2 = 10$ terms. Equating coefficients, we obtain ten equations. Since T is still arbitrary, these reduce to nine equations among the sixteen coefficients of the transformation. Hence there remain seven arbitrary constants.

It is remarkable that in spite of this *the most general rotation and expansion can be obtained by quaternion multiplication.* Let $\pi = \delta + i\alpha$

$+j\beta + k\gamma$ be another constant quaternion. Then we may show, just as before, that the transformation $q' = q \cdot \pi$, which differs from the preceding one only in that the order is reversed, represents a rotation and expansion of R_4. Hence the combined transformation

$$\text{(II)} \quad q' = p \cdot q \cdot \pi = (d + ia + jb + kc) \cdot q \cdot (\delta + i\alpha + j\beta + k\gamma)$$

also represents such a rotation and expansion. This transformation contains only seven (not eight) arbitrary constants, for the transformation remains unchanged if we multiply a, b, c, d by any real number and divide $\alpha, \beta, \gamma, \delta$ by the same number. It is therefore plausible that this combined transformation represents the general rotation and expansion of four-dimensional space. This beautiful result is actually true, as was shown by Cayley. I shall restrict myself to the mention of the historical fact, in order not to be drawn into too great detail. The formula is given in Cayley's paper *on the homographic transformation of a surface of the second order into itself* [1], in 1854, and also in certain other papers of his [2].

This formula of Cayley's has the great advantage that it enables us to grasp at once the combination of two rotations and expansions. Thus, if a second rotation and expansion be given by the equation

$$q'' = w'' + ix'' + jy'' + kz'' = p' \cdot q' \cdot \pi',$$

where p' and π' are new given quaternions, we find, by (II),

$$q'' = p' \cdot (p \cdot q \cdot \pi) \cdot \pi',$$

whence, by the associative law,

$$q'' = (p' \cdot p) \cdot q \cdot (\pi \cdot \pi')$$

or

$$q'' = r \cdot q \cdot \varrho$$

where $r = p' \cdot p$ and $\varrho = \pi \cdot \pi'$ are definite new quaternions. We have therefore obtained an expression for the rotation and expansion that carries q into q'' in precisely the old form, and we see that the multipliers which precede and follow q in the quaternion product are, respectively, the products of the corresponding multipliers of q in the separate transformations which were combined, the order of the factors being necessarily as shown in the formula.

This four-dimensional representation may seem unsatisfactory, and there may be a desire for something more tangible which can be represented in ordinary three-dimensional space. We shall therefore show that we can obtain similar formulas for the similar three-dimensional

---

[1] Journal für Mathematik, 1855. Reprinted in Cayley's Collected Papers, vol. 2, p. 133. Cambridge 1889.

[2] See, for example, *Recherches ultérieures sur les déterminants gauches*, loc. cit., p. 214.

operations by a simple specialization of the formulas just given. Indeed the importance of quaternion multiplication for ordinary physics and mechanics is based upon these very formulas. I have said "ordinary", because I do not desire at this point to explain those generalizations of these science for which the preceding formulas apply without any modification. These generalizations are more immediate, however, than you may suppose. The new developments of electrodynamics which are associated with the *principle of relativity*, are essentially nothing else than the logical use of rotations and expansions in a four-dimensional space. These ideas have been presented and enlarged upon recently by Minkowski[1].

Let us remain, however, in three-dimensional space. In such a space, a rotation and expansion carries a point (x, y, z) into a point (x', y', z') in such a way that

$$x'^2 + y'^2 + z'^2 = M^2(x^2 + y^2 + z^2),$$

where M denotes the ratio of expansion of every length. Since the general linear transformation of (x, y, z) into (x', y', z') contains nine coefficients, and since the left-hand side of the preceding equation, after the insertion of the values of x', y', z', becomes a quadratic form in x, y, z with six terms, the comparison of coefficients in the preceding equation leads to six equations, which reduce to five if the value of M is supposed arbitrary. Therefore the nine original coefficients of the linear transformation, which are subject to these five conditions, are reduced to four arbitrary constants. (Compare p. 67.) If such a linear transformation has a positive determinant, it represents, as was stated on p. 67, a rotation of space about the origin, together with an expansion in the ratio $1/M$. If the determinant is negative, however, the transformation represents a rotation and expansion, combined with a *reflection*, such as, for example, the reflection defined by the equations $x = -x', y = -y', z = -z'$. Moreover, it can be shown that the determinant of the transformation must have one of the two values $\pm M^3$.

In order to represent these relationships by means of quaternions, let us first reduce the variable quaternions q and q' to their vectorial parts:

$$q' = ix' + jy' + kz', \qquad q = ix + jy + kz,$$

which we shall think of as the three-dimensional vectors joining the origin to the positions of the point before and after the transformation, respectively. We shall show that *the general rotation and expansion*

[1] Since this was written, an extensive literature on the special theory of relativity mentioned above has appeared. Let me mention here my address *Über die geometrischen Grundlagen der Lorentzgruppe*, Jahresbericht der deutschen Mathematiker-Vereinigung, vol. 19 (1910), p. 299, reprinted in Klein's *Gesammelte mathematische Abhandlungen*, vol. 1, p. 533.

of the three-dimensional space is given by the formula (II) *if* p *and* π *have conjugate values, that is, if we write* $q' = p \cdot q \cdot \bar{p}$; *or, in expanded form,*

$$(1) \quad \begin{cases} ix' + jy' + kz' \\ \quad = (d + ia + jb + kc)\,(ix + jy + kz)\,(d - ia - jb - kc). \end{cases}$$

In order to prove this, we must show first that the scalar part of the product on the right vanishes; that is, that q' is indeed a *vector*. To do this, we first mutiply p by q according to the rule for quaternion multiplication, and we find

$$q' = [-ax - by - cz + i\,(dz + bz - cy) \\ + j(dy + cx - az) + k\,(dz + ay - bx)] \cdot [d - ia - jb - kc].$$

After another quaternion multiplication, we actually find the scalar part of q' to be zero, whereas we find for the components of the vector part the expressions

$$(2) \quad \begin{cases} x' = (d^2 + a^2 - b^2 - c^2)x + 2(ab - cd)y + 2(ac + bd)z \\ y' = 2(ba + cd)x + (d^2 + b^2 - c^2 - a^2)y + 2(bc - ad)z \\ z' = 2(ca - bd)x + 2(cb + ad)y + (d^2 + c^2 - a^2 - b^2)z \end{cases}$$

That these formulas actually represent a rotation and expansion becomes evident if we write the tensor equation for (1), which, by (I), is

$$x'^2 + y'^2 + z'^2 = (d^2 + a^2 + b^2 + c^2)\,(x^2 + y^2 + z^2)\,(d^2 + a^2 + b^2 + c^2),$$

or

$$x'^2 + y'^2 + z'^2 = T^4 \cdot (x^2 + y^2 + z^2),$$

where $T = \sqrt{d^2 + a^2 + b^2 + c^2}$ denotes the tensor of p. Hence, our transformation is precisely a rotation and expansion (see p. 69), provided the determinant is positive; otherwise it is such a transformation combined with a reflection. In any case, the ratio of expansion is $M = T^2$. As remarked above, the determinant must have one of the two values $\pm M^3 = \pm T^6$. If we consider the transformation for all possible values of the parameters a, b, c, d which correspond to the same tensor value T, which must obviously be different from zero, we see that the determinant must *always* have the value $+T^6$ if it has that value for any *single* system of values of a, b, c, d; for the determinant is a continuous function of a, b, c, d, and therefore it cannot suddenly change in value from $+T^6$ to $-T^6$ without taking on intermediate values. One set of values for which the determinant is positive is $a = b = c = 0$, $d = T$, since, by (2), the value of the determinant for these values of a, b, c, d, is

$$\begin{vmatrix} d^2, & 0, & 0 \\ 0, & d^2, & 0 \\ 0, & 0, & d^2 \end{vmatrix} = d^6 = +T^6.$$

It follows that the sign is always positive, and hence (1) always represents actually a rotation and expansion. It is easy to write down a transformation which combines a reflection with a rotation and an expansion, for we need only combine the preceding transformation with the reflection $x' = -x, y' = -y, z' = -z$, which is equivalent to writing the quaternion equation $\bar{q}' = p \cdot q \cdot \bar{p}$.

We shall now show that, conversely, every rotation and expansion may be written in the form (1), or in the equivalent form (2). In the first place, this formula contains the four arbitrary constants which, as we saw on p. 69, are necessary for the general case. That we can actually obtain any desired value of the expansion-ratio $M = T^2$, any desired position of the axis of rotation, and any desired angle of rotation, by a suitable choice of the four arbitrary constants, can be seen by means of the following formulas. Let ξ, η, ζ denote the direction cosines of the axis of rotation, and let ω denote the angle of rotation. We have, of course, the well known relation

$$\xi^2 + \eta^2 + \zeta^2 = 1. \tag{3}$$

I shall now prove that a, b, c, d are given by the equations

$$\left\{\begin{array}{l} d = T \cdot \cos\frac{\omega}{2}; \\ a = T \cdot \xi \cdot \sin\frac{\omega}{2}, \qquad b = T \cdot \eta \cdot \sin\frac{\omega}{2}, \qquad c = T \cdot \zeta \cdot \sin\frac{\omega}{2}, \end{array}\right. \tag{4}$$

which, by (3), obviously satisfy the condition

$$d^2 + a^2 + b^2 + c^2 = T^2.$$

When these relations have been proved, we can evidently obtain the correct values of a, b, c, d for any given values of $T, \xi, \eta, \zeta, \omega$.

To prove the relations (4), let us remark first that if a, b, c, d are given, the quantities ω, ξ, η, ζ are determined, and in such a way that (3) is satisfied. For, squaring and adding the equations (4), since T is the tensor of the quaternion $p = d + ia + jb + kc$, we have

$$1 = \cos^2\frac{\omega}{2} + \sin^2\frac{\omega}{2}(\xi^2 + \eta^2 + \zeta^2),$$

whence we see that (3) holds. It follows that ξ, η, ζ are fully determined by the relations

$$a : b : c = \xi : \eta : \zeta, \tag{4'}$$

which appear directly from (4). These equations express the fact that the point (a, b, c) lies on the axis of revolution of the transformation. This fact is easy to verify, for if we put $x = a, y = b, z = c$ in (2), we find

$$\begin{aligned} x' &= (d^2 + a^2 + b^2 + c^2)\, a = T^2 \cdot a, \\ y' &= (d^2 + a^2 + b^2 + c^2)\, b = T^2 \cdot b, \\ z' &= (d^2 + a^2 + b^2 + c^2)\, c = T^2 \cdot c \end{aligned}$$

that is, the point (a, b, c) remains on the same ray through the origin, which identifies it as a point on the axis of revolution. It remains ɔnly to prove that the angle ω defined by (4) is actually the angle of rotation. This demonstration requires extended discussion which [can avoid now by remarking that the transformation (2) for $T = 1$ reduces precisely to the transformation given by Euler for the revolution ɔf the axes through the angle ω about an axis of revolution whose lirection cosines are ξ, η, ζ. This is to be found, for example, in Klein-Sommerfeld, *Theorie des Kreisels*, volume 1[1], where explicit mention ɔf the theory of quaternions is given, or in Baltzer, *Theorie und Anwendung der Determinanten*[2].

Finally, if we substitute the values given by (4) in the equation (1), ve obtain the very brief and convenient equation in quaternion form or the revolution through an angle ω about an axis whose direction osines are ξ, η, ζ, combined with an expansion of ratio T^2:

$$\text{(5)}\quad \begin{cases} ix' + jy' + kz' = T^2\left\{\cos\frac{\omega}{2} + \sin\frac{\omega}{2}(i\xi + j\eta + k\zeta)\right\}\cdot\{ix + jy + kz\} \\ \qquad\qquad\cdot\left\{\cos\frac{\omega}{2} - \sin\frac{\omega}{2}(i\xi + j\eta + k\zeta)\right\}. \end{cases}$$

his formula expresses in a form that is easy to remember Euler's ɔrmulas for rotation: the multipliers which precede and follow the ector $ix + jy + kz$, are, respectively, the two conjugate quaternions hose tensor is unity (so-called *versor*, that is, "rotator", in contra-istinction to *tensor*, "stretcher"), and then the whole result is to be ıultiplied by a scalar factor which is the expansion-ratio.

We shall proceed now to show that when we specialize these formulas :ill further to two-dimensions, they become the well known formulas ɔr the representation of a rotation and expansion of the xy plane by ıeans of the multiplication of two complex numbers. (See p. 57.) or this purpose, let us choose the axis of rotation as the z axis $= \eta = 0, \zeta = 1)$. Then the formula (5), for $z = z' = 0$, may be ritten in the form

$$ix' + jy' = T^2\left(\cos\frac{\omega}{2} + k\sin\frac{\omega}{2}\right)(ix + jy)\left(\cos\frac{\omega}{2} - k\sin\frac{\omega}{2}\right),$$

, upon multiplication with due regard to the rules for products of the ıits,

$$\begin{aligned} {}' + jy' &= T^2\left\{\cos\frac{\omega}{2}(ix + jy) + \sin\frac{\omega}{2}(jx - iy)\right\}\left\{\cos\frac{\omega}{2} - k\sin\frac{\omega}{2}\right\} \\ &= T^2\left\{\cos^2\frac{\omega}{2}(ix + jy) + 2\sin\frac{\omega}{2}\cos\frac{\omega}{2}(jx - iy) - \sin^2\frac{\omega}{2}(ix + jy)\right\} \\ &= T^2\{(ix + jy)\cos\omega + (jx - iy)\sin\omega\} \\ &= T^2(\cos\omega + k\sin\omega)(ix + jy). \end{aligned}$$

[1] Leipzig 1897; 2nd printing, 1914. [2] Fifth edition, Leipzig 1881.

If we now multiply both sides by the right-hand factor $(-i)$, we obtain

$$x' + ky' = T^2 (\cos\omega + k \sin\omega)\,(x + ky),$$

which is precisely the rule for multiplying two ordinary complex numbers, and which can be interpreted as a rotation through an angle ω, together with an expansion in the ratio T^2, except that we have used the letter k in place of the usual letter i to denote the imaginary unit $\sqrt{-1}$.

Let us now return to three-dimensional space, and let us modify the formula (1) so that it shall represent a pure rotation without an expansion. To do so, we must replace x', y', z' by $x' \cdot T^2, y' \cdot T^2, z' \cdot T^2$, that is, we must replace q' by $q' \cdot T^2$. If we notice that $p^{-1} = 1/p = p/T^2$, we may write the formula for a pure rotation in the form

$$ix' + jy' + kz' = p \cdot (ix + jy + kz) \cdot p^{-1}. \tag{6}$$

There is no loss of generality if we assume that p is a quaternion whose tensor is unity, that is,

$$p = \cos\frac{\omega}{2} + \sin\frac{\omega}{2}\,(i\xi + j\eta + k\zeta), \quad \text{where} \quad \xi^2 + \eta^2 + \zeta^2 = 1,$$

whence we see that (6) results from (5) if T is set equal to unity. The formula was first stated in this form by Cayley in 1845[1].

We may express the composition of two rotations in a particularly simple form, precisely as we did above for four-dimensional space. Given a second rotation

$$ix'' + jy'' + kz'' = p'\,(ix' + jy' + kz')\,p'^{-1},$$

where

$$p' = \cos\frac{\omega'}{2} + \sin\frac{\omega'}{2}\,(i\xi' + j\eta' + k\zeta')$$

the direction cosines of the axis of rotation being ξ', η', ζ', and the angle of rotation being ω', we may write

$$ix'' + jy'' + kz'' = p' \cdot p \cdot (ix + jy + kz) \cdot p^{-1} \cdot p'^{-1}$$

as the equation for the resultant rotation. Hence the direction cosines of the axis or rotation, ξ'', η'', ζ'', and the angle of rotation, ω'', for the resultant rotation, are given by the equation

$$p'' = \cos\frac{\omega''}{2} + \sin\frac{\omega''}{2}\,(i\xi'' + j\eta'' + k\zeta'') = p' \cdot p.$$

We have therefore found a brief and simple expression for the composition of two rotations about the origin, whereas the ordinary formulas for expressing the resultant rotation appear rather complicated. Since any quaternion may be expressed as the product of a real number

[1] *On certain results relating to quaternions*, Collected Mathematical Papers, vol. 1 (1889), p. 123. According to Cayley's own statement (vol. 1, p. 586), however, Hamilton had discovered the same formula independently.

(its tensor) and the versor of a rotation, we have also found a simple geometric interpretation of quaternion multiplication as the composition of the rotations. The fact that quaternion multiplication is not commutative then corresponds to the well known fact that the order of two rotations about a point cannot be interchanged, in general, without changing the result.

If you desire to make a study of the historical development of the representations and applications of quaternions which we have discussed, I would recommend to you an extremely valuable report on dynamics written by Cayley himself: *Report on the progress of the solution of certain special problems of dynamics*[1].

I shall close with certain general remarks on the value and the dissemination of quaternions. For such a purpose, one should distinguish between the general quaternion calculus and the simple rule for quaternion multiplication. The latter, at least, is certainly of very great usefulness, as appears sufficiently from the preceding discussion. The general quaternion calculus, on the other hand, as Hamilton conceived it, embraced addition, multiplication, and division of quaternions, carried to an arbitrary number of steps. Thus Hamilton studied the algebra of quaternions; and, since he investigated also infinite processes, he may be said to have created a quaternion theory of functions. Since the commutative law does not hold, such a theory takes on a totally different aspect from the theory of ordinary complex variables. It is just to say, however, that these general and far-reaching ideas of Hamilton have not justified themselves, for there have not arisen any vital relationships and interdependencies with other branches of mathematics and its applications. For this reason, the general theory has aroused little general interest.

It is in mathematics, however, as it is in other human affairs: there are those whose views are calmly objective; but there are always some who form regrettable personal prejudices. Thus the theory of quaternions has enthusiastic supporters and bitter opponents. The supporters, who are to be found chiefly in England and in America, adopted in 1907 the modern plan by founding an "Association for the Promotion of the Study of Quaternions". This organization was established as a thoroughly international institution by the Japanese mathematician Kimura, who had studied in America. Sir Robert Ball was for some time its president. They foresaw great possible developments of mathematics to be secured through intensive study of quaternions. On the other hand, there are those who refuse to listen to anything about quaternions, and who go so far as to refuse to consider the very useful idea of quaternion mul-

[1] Report of the British Association for the Advancement of Science, 1862; reprinted in Cayley's Collected Mathematical Papers, Cambridge, vol. 4 (1891), pp. 552ff.

tiplication. According to the view of such persons, all computation with quaternions amounts to nothing but computation with the four components; the units and the multiplication table appear to them to be superfluous luxuries. Between these two extremes, there are many who hold that we should always distinguish carefully between scalars and vectors.

4. Complex Numbers in School Instruction

I shall now leave the theory of quaternions and close this chapter with some remarks about the role which these concepts play in the curriculum of the schools. No one would ever think of bringing up quaternions in a secondary school, but *the common complex numbers* $x + iy$ *always come up for discussion.* Perhaps it will be more interesting if, instead of telling you at length how it is done and how it ought to be done, I exhibit to you, by means of three books from different periods, *how instruction has developed historically.*

I put before you, first, a book by Kästner who had a leading position in Göttingen in the second half of the eighteenth century. In those days one still studied, at the university, those elementary mathematical things which later, in the thirties of the nineteenth century, went over to the schools. Accordingly, Kästner also gave lectures on elementary mathematics, which were heard by large numbers of non-mathematical students. His book, which formed the basis of these lectures, was called *Mathematische Anfangsgründe**. The portion which interests us here is the second division of the third part: *Anfangsgründe der Analysis endlicher Größen***[1]. The treatment of imaginary quantities begins there on page 20 in something like the following words: "Whoever demands the extraction of an even root of a 'denied' quantity (one said 'denied', then, instead of 'negative'), demands an impossibility, for there is no 'denied' quantity which would be such a power". This is, in fact, quite correct. But on page 34 one finds: "Such roots *are called* impossible or imaginary", and, without much investigation as to justification, one proceeds quietly to operate with them as with ordinary numbers, notwithstanding their existence has just been disputed—as though, so to speak, the meaningless became suddenly usable through receiving a name. You recognize here a reflex of Leibniz's point of view, according to which, imaginary numbers were really something quite foolish but they led, nevertheless, in some incomprehensible way, to useful results.

Kästner was, moreover, a stimulating writer; he achieved quite a place in the literature as a coiner of epigrams. To cite only one of many examples, he expatiates, in the introduction of this book mentioned

[1] Third edition. Göttingen 1794.

* *Elements of Mathematics.*

** *Elemements of Analysis of Finite Quantities.*

above, on the *origin of the word algebra*, which, indeed, as the article "al" indicates, comes from the Arabic. According to Kästner, an algebraist is a man who "makes" fractions "whole", who, that is, treats rational functions and reduces them to a common denominator, etc. It is said to have referred, originally, to the practice of a surgeon in mending broken bones. Kästner then cites Don Quixote, who went to an algebraist to get his broken ribs set. Of course, I shall leave undecided, whether Cervantes really adopted this form of expression or whether this is only a lampoon.

The second work which I put before you is more recent, by a whole series of years, and comes from the Berlin professor M. Ohm: *Versuch eines vollständig konsequenten Systems der Mathematik**[1]; a book with purpose similar to that of Kästner and at one time widely used. But Ohm is much nearer the modern point of view, in that he speaks clearly of the *principle of the extension of the number system*. He says, for example, that, just like negative numbers, so $\sqrt{-1}$ must be added to the real numbers as a *new thing*. But even his book lacks a geometric interpretation, since it appeared before the epoch-making publication by Gauss (1831).

Finally, I lay before you, out of the long list of modern school books, one that is widely used: *Bardeys Aufgabensammlung*[2]. The *principle of extension* comes to the fore here, and, in due course, the *geometric interpretation* is explained. This may be taken as the general position of school instruction today, even if, at isolated places, the development has remained at the earlier level. The point of view adopted in this book seems to me to yield the treatment best adapted to the schools. Withhout tiring the pupil with a systematic development, and without, of course, going into logically abstract explanations, one should explain *complex numbers as an extension of the familiar number concept*, and should avoid any touch of mystery. Above all, one should accustom the pupil, at once, to the *graphic geometric interpretation in the complex plane!*

With this, we come to the end of the first main part of the course, which was dedicated to arithmetic. Before going over to similar discussions of algebra and analysis, I should like to insert a somewhat extended historical appendix in order to throw new light upon the general conduct of instruction at present, and upon those features of it which we would improve.

---

[1] Nine volumes. Berlin 1828. Vol. I: *Arithmetik und Algebra*, p. 276.

* *An Attempt to Construct a Consistent System of Mathematics.*

[[2] See also the *Reformausgabe* of Bardeys *Aufgabensammlung*, revised by W. Lietzmann and P. Zülke. Oberstufe. Verlag Teubner. Leipzig.]—See also Fine, H., *The Number-System in Algebra*. Heath. Fine, H., College Algebra. Ginn.

Concerning the Modern Development and the General Structure of Mathematics

Let me proceed from the remark that, in the *history of the development* of mathematics up to the present time, we can distinguish clearly *two different processes of growth*, which now change places, now run side by side independent of one another, now finally mingle. It is difficult to put into vivid language the difference which I have in mind, because none of the current divisions fits the case. You will, however, understand from a concrete example, namely, if I show how one would compile the *elementary chapters of the system of analysis* in the sense of each of these two processes of development.

If we follow the one process, which we will call briefly *Plan A*, the following system presents itself, the one which is most widespread in the schools and in elementary textbooks.

1. At the head stands the *formal theory of equations*, that is to say, the *operating with rational integral functions* and the handling of the cases in which *algebraic equations can be solved by radicals*.

2. The *systematic pursuit of the idea of power and its inverses* yields *logarithms*, which prove to be so useful in numerical calculation.

3. Whereas (up to this point) the analytic development is kept quite separate from geometry, one now borrows from this field, which yields the *definitions of a second kind of transcendental functions, the trigonometric functions*, the further theory of which is built up as a *new separate subject*.

4. Then follows the so called "algebraic analysis", which teaches *the development of the simplest functions into infinite series*. One considers the *general binomial*, the *logarithm* and its inverse, the *exponential function*, together with *the trigonometric functions*. Similarly, the *general theory of infinite series and of operations with them* belongs here. It is here that the *surprising relations between the elementary transcendentals* appear, in particular the famous *Euler formula*

$$e^{ix} = \cos x + i \sin x .$$

Such relations seem the more remarkable because the functions which occur in them had been originally defined in entirely separate fields.

5. The consistent continuation, beyond the schools, of this structure, is the *Weierstrass theory of functions of a complex variable*, which begins with the properties of *power series*.

Let us now set over against this, in condensed form the *second process of development*, which I shall call *Plan B*. Here the controlling *thought is that of analytic geometry*, which seeks a *fusion of the perception of number with that of space*.

1. We begin with the *graphical representation of the simplest functions*, of polynomials, and rational functions of one variable. The point in

which the curves so obtained meet the axis of abscissas put in evidence the *zeros of the polynomials*, and this leads naturally to the *theory of the approximate numerical solution of equations*.

2. The *geometric picture of the curve supplies naturally the intuitive source both for the idea of the differential quotient and that of the integral.* One is led to the former by the *slope of the curve*, to the latter by the *area which is bounded by the curve and the axis of abscissas*.

3. In all those cases in which the *integration process* (or the process of quadrature, in the proper sense of that word) cannot be carried out explicitly with rational and algebraic functions, the process itself gives *rise to new functions*, which are thus introduced in a thoroughly natural and uniform manner. Thus, the *quadrature of the hyperbola* defines the *logarithm*

$$\int_1^x \frac{dx}{x} = \log x,$$

while the *quadrature of the circle* can easily be reduced to the integral

$$\int_0^x \frac{dx}{\sqrt{1-x^2}} = \operatorname{arc}\sin x,$$

that is, to the *inverses of the trigonometric functions*. You know that the same line of thought, pursued farther, leads to new classes of functions of higher order, in particular to *elliptic functions*.

4. The *development into infinite power series of the functions thus introduced* is obtained by means of a *uniform principle*, namely *Taylor's theorem*.

5. This method carried higher, yields the *Cauchy-Riemann theory of analytic functions of a complex variable*, which is built upon the *Cauchy-Riemann differential equations* and the *Cauchy integral theorem*. If we try to put the *result of this survey* into definite words, we might say *that Plan A is based upon a more particularistic conception of science which divides the total field into a series of mutually separated parts and attempts to develop each part for itself, with a minimum of resources and with all possible avoidance of borrowing from neighboring fields.* Its ideal is to *crystallize out each of the partial fields into a logically closed system.* On the contrary, *the supporter of Plan B lays the chief stress upon the organic combination of the partial fields, and upon the stimulation which these exert one upon another. He prefers, therefore, the methods which open for him an understanding of several fields under a uniform point of view. His* ideal is the *comprehension of the sum total of mathematical science as a great connected whole.*

One cannot well be in doubt as to which of these two methods has more life in it, as to which would grip the pupil more, in so far as he is not endowed with a specific abstract mathematical gift. In order to bring this home, think only of the *example of the functions* e^x *and* $\sin x$,

about which we shall later have much to say along just this line! In Plan A, which the schools, unfortunately, follow almost exclusively, both functions come up in thoroughly heterogeneous fashion: e^x or, as the case may be, the *logarithm*, is introduced as a *convenient aid in numerical calculation*, but $\sin x$ appears in the *geometry of the triangle*. How can one understand, thus, that the two are so simply connected, and, more, that the two appear again and again in the most widely differing fields which have not the least to do, either with the technique of numerical calculation or with geometry, and always of their own accord, as the natural expression of the laws that govern the subject under discussion? How far these possibilities of application go is shown by the names *compound interest law* or *law of organic growth*, which have been applied to e^x, and likewise by the fact that $\sin x$ plays a central role wherever one has to do with *vibrations*. *But in Plan B these connections make their appearance quite intelligibly, and in accord with the significance of the functions, which is emphasized from the start.* The functions e^x and $\sin x$ arise here, indeed, from the same source, the quadrature of simple curves, and one is soon led from there, as we shall see later on, to the *differential equations of simplest type*

$$\frac{d e^x}{d x} = e^x, \qquad \frac{d^2 \sin x}{d x^2} = -\sin x,$$

respectively, which lie naturally at the basis of all those applications.

For a complete understanding of the development of mathematics we must, however, think of still a *third Plan C*, which, along side of and within the processes of development A and B, often plays an important role. It has to do with a method which one denotes by the word *algorithm*, derived from a mutilated form of the name of an Arabian mathematician. *All ordered formal calculation* is, at bottom, algorithmic, in particular, the *calculation with letters* is an algorithm. We have repeatedly emphasized what an important part in the development of the science has been played by the algorithmic process, as a *quasi-independent, onward-driving force, inherent in the formulas*, operating apart from the intention and insight of the mathematician, at the time, often indeed in opposition to them. In the beginnings of the infinitesmal calculus, as we shall see later on, the algorithm has often forced new notions and operations, even before one could justify their admissibility. Even at higher levels of the development, these algorithmic considerations can be, and actually have been, very fruitful, so that one can justly call them the *groundwork of mathematical development*. We must then completely ignore history, if, as is sometimes done today, we cast these circumstances contemptuously aside as mere "formal" developments.

Let me now follow more carefully through the history of mathematics the contrast of these different directions of work, confining myself of course

to the most important features of the development. The *thoroughgoing difference between A and B, within the whole field of mathematics,* will appear here more clearly than it did above, where our thoughts were directed only to analysis.

With the *ancient Greeks* we find a *sharp separation between pure and applied mathematics,* which goes back to Plato and Aristotle. Above all, the well known *Euclidean structure of geometry* belongs to pure mathematics. In the applied field they developed, especially, *numerical calculation,* the so called *logistics* (λόγος = general number, see p. 32). To be sure, the logistics was not highly regarded, and you know that this prejudice has, to a considerable extent, maintained itself to this day —mainly, it is true, with only those persons who themselves cannot calculate numerically. The slight esteem for logistics may have been due in particular to its having been developed in connection with *trigonometry* and the needs of *practical surveying,* which to some does not seem sufficiently aristocratic. In spite of this fact, it may have been raised somewhat in general esteem by its application in astronomy, which, although related to geodesy, always has been considered one of the most aristocratic fields. You see, even with these few remarks, that the Greek cultivation of science, with its sharp separation of the different fields, each of which was represented with its rigid logical articulation, *belonged entirely in the plan of development A.* Nevertheless the *Greeks were not entire strangers to reflections in the sense of Plan B,* and these may have served them for heuristic purposes, and for a first communication of their discoveries, even if the form *A* appeared to them indispensable for the final presentation. This is indicated quite pointedly in the recently discovered *manuscript of Archimedes*[1], in which he exhibits his calculations of volume through mechanical considerations, in a thoroughly modern, pleasing way, which has nothing to do with the rigid Euclidean system.

Besides the Greeks, in ancient times, the *Hindus,* especially, played a mathematical role as *creators of our modern system of numerals,* and later the *Arabs, as its transmitters.* The *first beginnings of operating with letters* were made also by the Hindus. These advances belong obviously to the *algorithmic course of development C.*

Coming now to *modern times,* we can, first of all, *date the mathematical renaissance from about 1500,* which produced an entire *series of great discoveries.* As an example, I mention the *formal solution of the cubic equation* (Cardan's formula), which was contained in the *"Ars Magna" of Cardano,* published in 1545, in Nürnberg. This was a most significant work, which holds the germs of the modern algebra, reaching out beyond

[1] Cf. Heiberg und Zeuthen, *Eine neue Schrift des Archimedes.* Leipzig 1907. Reprint from Bibliotheca Mathematica. Third series, vol. VIII. See also Heath, *T. L., The Works of Archimedes.* Cambridge University Press.

the scheme of ancient mathematics. To be sure, this work is not Cardano's own, for he is said to have taken from other authors not merely his famous formula but other things as well.

After 1550 *trigonometric calculation* was in the foreground. *The first great trigonometric tables* appeared in response to the needs of *astronomy*, in connection with which I will mention only the name of Copernicus. *From about 1600 on*, the *invention of logarithms* continued this development. The first logarithmic tables, which a Scotchman Napier (or Nepér) compiled, contained, in fact, only the logarithms of trigonometric functions. Thus we see, during these hundred years, a path of development which corresponds to the *Plan A*.

We come now, in the *seventheenth century*, to the *modern era proper*, in which the *Plan B comes distinctly into the foreground*. In *1637* appeared the *analytic geometry of Descartes*, which supplies the fundamental connection between number and space for all that follows. A reprint[1] makes this work conveniently accessible. Now come, in close sequence with this, the *two great problems of the seventeenth century*, the *problem of the tangent*, and *the problem of quadrature*, in other words, the *problems of differentiating and integrating*. For the development of differential and integral calculus, in a proper sense, there was lacking only the knowledge that these two problems are closely connected, *that one is the inverse of the other*. A recognition of this fact was the *principal item in the great advance* which was made at the end of the seventeenth century.

But before this, in the course of the century, the *theory of infinite series, in particular, of power series*, made its appearance, and not, indeed, as an independent subject, in the sense of the algebraic analysis of today, but in *closest connection with the problem of quadrature*. Nicolaus Mercator (the German name "Kaufmann" latinized; 1620—1687), not the inventor of the Mercator projection, was a pioneer here. He had the keen idea of converting the fraction $1/(1+x)$ into a series, by dividing out, and of *integrating this series term by term*, in order to get the *series development for* $\log(1+x)$:

$$\log(1+x) = \int_0^x \frac{dx}{1+x} = \int_0^x (1 - x + x^2 - + \cdots)\, dx = x - \frac{x^2}{2} + \frac{x^3}{3} - + \cdots$$

That is the substance of his procedure, although he did not, of course, use our simple symbols $\int$, $dx, \ldots$, but rather a much more clumsy form of expression. In the sixties, Isaac Newton (1643—1727) took over this process, to apply it to the *series for the general binomial*, which he had set up. In this process he drew his *conclusions by analogy*, basing

[1] Descartes, R., *La Géométrie*. Nouvelle édition. Paris 1886. Translation by Smith, D. E., and Latham, M. L., 1925. Open Court.

them on the known simplest cases, without having a rigorous proof and without knowing the limits within which the series development was valid. We observe here, again, the operation of the *algorithmic process C*. By applying the binomial series to $\frac{1}{\sqrt{1-x^2}} = (1-x^2)^{-1/2}$ and using Mercator's process, he gets the *series for* $\int_0^x \frac{dx}{\sqrt{1-x^2}} = \text{arc } \sin x$. By a very skillful *inversion of this series*, and also of the one for $\log x$, he finds the series for $\sin x$ and for e^x. The conclusion of this chain of discoveries is due to Brook Taylor (1685—1731) who, in 1715, published his *general principle for developing functions into power series.*

As is indicated above, the *origin of infinitesimal calculus, at the end of the seventeenth century,* was due to G. W. Leibniz (1646—1716) and Newton. The fundamental idea with Newton is the *notion of flowing*. Both variables x, y, are tought of as functions, $\varphi(t)$, $\psi(t)$, of the time t; and as the time "flows", they flow also. Newton, accordingly, calls the variable *fluens* and designates as *fluxion* $\dot{x}$, $\dot{y}$, that which we call differential coefficient. You see how everything here is *based firmly on intuition.*

It was much the same with the representation of Leibniz, whose first publication appeared in 1684. He himself declares that his greatest discovery was the *principle of continuity in all natural phenomena*, that "Natura non facit saltum". He bases his mathematical developments upon this concept, another example of the *Plan B*. However, the *influence of the algorithm C is very strong*, also, with Leibniz. We get from him the algorithmically valuable symbols dy/dx and $\int f(x)\,dx$.

The sum total, however, of this cursory view is that the *great discoveries of the seventeenth century belong primarily to the plan of development B*.

In the eighteenth century, this period of discovery continues at first *in the same direction*. The most distinguished names to be mentioned here are L. Euler (1707—1783) and J. L. Lagrange (1736—1813). Thus the *theory of differential equations, in the most general sense, including the calculus of variations*, were developed, and *analytical geometry and analytical mechanics were extended*. Everywhere there was a gratifying advance, just as in geography, after the discovery of America, the new countries were first traversed and explored in all directions. But just as there was, as yet, no thought of exact surveys, just as at first one had entirely false notions as to the location of these new places (Columbus, indeed, thought at first, that he had reached Eastern Asia!), just so, in the newly conquered region of mathematics, that of infinitesimal calculus, one was, at first, far removed from a reliable logical orientation. Indeed one even cherished illusions concerning the relation of the calculus to the older familiar fields, in that one looked upon infinitesimal calculus as something *mystical* that in no way submitted to a logical analysis.

Just how untrustworthy the ground was on which the theory stood, became manifest only when it was attempted to prepare textbooks which should present the new subject in an intelligible way. Then it became evident that the method of procedure B was no longer adequate, and it was Euler who first abandoned it. He had, to be sure, no serious doubts concerning infinitesimal calculus, but he thought that it caused too many difficulties and misgivings for the beginner. For this pedagogical reason he thought it advisable to give a preparatory course, such as he provided in his text book *Introductio in analysin infinitorum* (1748), and which we call today *algebraic analysis*. To this he relegated, in particular, the *theory of infinite series and other infinite processes*, which he then afterwards used as a foundation in constructing the infinitesimal calculus.

Lagrange took a much more radical course, nearly fifty years later, in his *Thèorie des fonctions analytiques, in 1797*. He could satisfy his scruples as to the current foundations of infinitesimal calculus only by discarding it entirely, as a general branch of knowledge, and by considering it as an aggregate of formal rules applying to certain special classes of functions. Indeed, he considers *exclusively such functions as can be expressed by means of power series:*

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots.$$

He calls these *analytic functions*, meaning thereby functions which appear in analysis and with which one can reasonably hope to do something. The *differential quotient of such a function, $f(x)$, is then defined, purely formally* , *by means of a second power series*, as we shall see later. Differential and integral calculus was concerned, then, with the mutual relations of power series. This restriction to formal consideration obviated, for a time, of course, a number of difficulties.

As you see, *the turn which Euler gave, and still more, the entire method of Lagrange, belongs strictly to the direction A, in that the perceptual genetic development is replaced by a rigorous closed circle of reasoning.* These two investigators have *had a profound influence upon instruction in the secondary schools*, and when the schools today study infinite series, or solve equations by means of power series according to the so called *method of indeterminate coefficients*, but decline to take up differential and integral calculus proper, they are *exhibiting precisely the after effect of Euler's "introductio" and of Lagrange's thought.*

The *nineteenth century*, to which we come now, begins primarily with a *more secure foundation of higher analysis, by means of criteria of convergence*, about which one had hitherto thought but little. The eighteenth century was the "blissful" period, during which one did not distinguish between good and bad, convergent and divergent. Even in Euler's *Introductio*, divergent and convergent series appear peaceably

side by side. But, at the beginning of the new century Gauss (1777—1855) and Abel (1802—1829) made the first rigorous investigations regarding convergence; and in the twenties Cauchy (1789—1857) developed, in lectures and in books, the *first rigorous founding of infinitesimal calculus in the modern sense.* He not only gives an exact *definition of the differential quotient, and of the integral, by means of the limit of a finite quotient and of a finite sum, respectively,* as had previously been done, at times; but, by means of the *mean-value theorem* he erects upon this, for the first time, a *consistent structure for the infinitesimal calculus.* We shall come back to this fully later on. These theories also partake of the nature of *Plan A*, since they work over the field in a logical sytematic way, quite apart from other branches of knowledge. *Meanwhile they had no influence upon the schools,* although they were thoroughly adapted to dispel the old prejudice against differential and integral calculus.

I shall now emphasize only a very little of the *further development of the nineteenth century.* In the first place, I shall speak of a few advances which lie in the *direction B: modern geometry, mathematical physics,* along with *theory of functions of a complex variable, according to Cauchy and Riemann.* The leaders, in the first working over of these three great fileds, were *the French.* This is the place to say a word, also, about the *style of mathematical presentation.* In Euclid, one finds everything according to the scheme "hypothesis, conclusion, proof", to which is added, sometimes, the "discussion", i. e., the determination of the limits which the considerations are valid. The belief is widespread that mathematics always moves thus four steps at a time. But just in the period of which we are speaking, there arose, especially among the French, a *new art form in mathematical presentation,* which might be called *artistically articulated deduction.* The works of Monge or, to mention a more recent book, the *Traité d'Analyse,* by Picard, read just like a well written gripping novel. This is the *style which fits the method of thought B,* whereas the *Euclidean presentation is related, in essence, to the method A.*

Of Germans who achieved distinction in these fields I should mention Jacobi (1804—1851), Riemann (1826—1866), and, coming to a somewhat later time, Clebsch (1833—1872), and the Norwegian Lie (1842—1899). These all belong essentially to the *direction B,* except that occasionally *an algorithmic touch* is noticeable with them.

From the middle of the century on, the *method of thought A* comes again to the front with Weierstrass (1815—1897). His activity, as teacher in Berlin, began in 1856. I have already instanced *Weierstrass function theory* as an example of *A*. The *more recent investigations concerning the axioms of geometry* belong, likewise, to the *type A*. One is concerned here with studies entirely in the Euclidean direction, which approach it, also, in the manner of presentation.

With this I bring our brief historical résumé to an end. Many points of view which could only be alluded to here will be brought up later for more complete discussion. As a summary, we might say that, *in the history of mathematics during the last centuries, both of our chief methods of investigation were of importance; that each of them, and sometimes the two in succession, have resulted in important advances of the science.* It is certain that mathematics will be able to advance uniformally in all directions, only if *neither of the two methods of investigation is neglected.* May each mathematician work in the direction which appeals to him most strongly.

Instruction in the secondary schools, however, as I have already indicated, has long been *under the one-sided control of the Plan A.* Any movement toward reform of mathematical teaching must, therefore, press for *more emphasis upon direction B.* In this connection I am thinking, above all, of an impregnation with the *genetic method of teaching,* of a stronger emphasis upon *space perception,* as such, and, particularly, of giving *prominence to the notion of function,* under *fusion of space perception and number perception!* It is my aim that these lectures shall serve this tendency, especially since these elementary mathematical books to which we are in the habit of going for advice, e g., those of Weber-Wellstein, Tropfke, M. Simon, represent the direction A almost exclusively. I called your attention, in the introduction, to this one-sidedness.

And now, gentlemen, enough of these diversions; let us pass to the next main subdivision of this course.

Part II

Algebra

Let me commence by mentioning a *few textbooks of algebra*, in order to introduce you somewhat to a very extensive literature. I suggest, first, Serret's *Cours d'algèbre*[1] which was much used in Germany, formerly, and had great merit. Now, however, we have two great widely used German textbooks: H. Weber's *Lehrbuch der Algebra*[2] and E. Netto's *Vorlesungen über Algebra*[3], each in two volumes; both treat with great fullness the most difficult parts of algebra and are well adapted for extensive special study; they seem to me to be too comprehensive for the average needs of prospective teachers and also too expensive. More fitting in the latter respect is the handy *Vorlesungen über Algebra*[4] by G. Bauer, which hardly goes beyond what the teacher should master[5]. On the *practical side*, for the numerical solution of equations, this book is supplemented by the little book *Praxis der Gleichungen* by C. Runge[6], which I can highly reccomend.

Turning now to the narrower subject, let me remark that I *cannot*, in the limits of this course of lectures, *give a systematic presentation of algebra*; I can give, rather, only a *one sided selection*, and it will be most fitting if I emphasize those things which are, unfortunately, neglected elsewhere, and which are calculated nevertheless to throw light upon school instruction. All of my algebraic developments will group themselves about *one* point, namely, about the *application to the solution of equations of graphical and, generally speaking, of geometrically perceptual methods*. This field alone is a very extensive and widely related chapter of algebra. Even from it, it is obviously possible to select only the most

[1] Third edition. Paris 1866 [sixth edition, 1910].

[2] Second edition. Braunschweig 1898/99. [New revision by R. Fricke. Vol. I. 1924.]

[3] Leipzig 1896/99. See also: Chrystal, *Textbook in Algebra* (2 volumes). Macmillan. Bôcher, M., *Introduction to Higher Algebra*. Macmillan.

[[4] Second edition. Leipzig 1910.]

[5] See also: Netto, E., *Elementare Algebra*, akademische Vorlesungen für Studierende der ersten Semester. [Second edition. Leipzig 1913, and H. Weber, *Lehrbuch der Algebra*. Small edition in one volume. Second printing. Braunschweig 1921.] See also: Fine, H., *College Algebra*. Ginn. Hall und Knight, *Higher Algebra*. Macmilian.

[[6] Second edition. Leipzig 1921.] See also: v. Sanden, H., *Practical Mathematical Analysis*. Dutton & Co.

important and interesting things; in doing this we shall come into organic relation with the most widely differing fields, so that we shall be *studying mathematics quite in the spirit of our system B.* In the first place, we shall treat equations in real unknowns in order that we may follow, later, with the consideration of complex quantities.

I. Real Equations with Real Unknowns

1. Equations with one parameter

We begin with a very simple case, which is susceptible of geometric treatment, namely with a real algebraic equation for the unknown x, in which a parameter λ appears:

$$f(x, \lambda) = 0.$$

We shall obtain a geometric representation most simply if we replace λ by a second variable y and think of

$$f(x, y) = 0$$

as a *curve in the xy plane* (see Fig. 19). *The points of intersection of this curve with the line $y = \lambda$, parallel to the x-axis, give the real roots of the equation $f(x, \lambda) = 0$.* When we have drawn the curve approximately, as we can easily do if f is not too complicated, we can see at a glance by displacing the parallel as λ varies, how the number of real roots changes. This plan is especially effective when *f is linear in λ,* i.e. with equations of the form

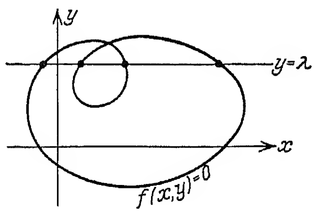


Fig. 19.

$$\varphi(x) - \lambda\psi(x) = 0$$

If φ and ψ are rational, the curve $y = \varphi(x)/\psi(x)$ will also be rational, and is easy to draw. In these cases one can often use this method to advantage in calculating approximately the roots of equations.

As an *example* consider *the quadratic equation*

$$x^2 + ax - \lambda = 0.$$

The curve $y = x^2 + ax$ is a *parabola,* and one can see at once for what values of λ the equations has two, one, or no real roots according as the horizontal line cuts the parabola in two, one, or no points (see Fig. 20). It seems to me that the presentation of such a simple and obvious construction would be very appropriate in the upper school classes.

As a second example let us take the *cubic equation*

$$x^3 + ax^2 + bx - \lambda = 0,$$

which gives us the *cubical parabola* $y = x^3 + ax^2 + bx$, whose appearance is different according to the values of a, b. In Fig. 21, it is assumed

that $x^2 + ax + b = 0$ has two real roots. It is easy to see how the parallels group themselves into those which intersect the curve in one

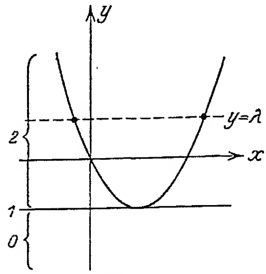


Fig. 20.

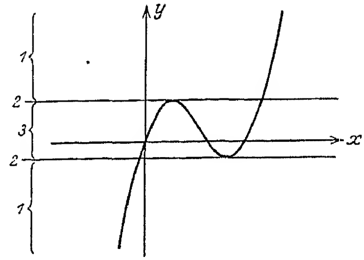


Fig. 21.

point and those which meet it in three; there can be two limiting positions which yield double roots.

2. Equations with two parameters

When several parameters, let us say two, appear in an equation, more skill is required to handle the problem graphically, but the results are more extensive and interesting. We shall limit ourselves to the case where the *two parameters* λ, μ *appear linearly*, and we shall write t for the unknown in the equation. The problem is to determine the real roots of the equation

$$\varphi(t) + \lambda \cdot \chi(t) + \mu \cdot \psi(t) = 0, \tag{1}$$

where φ, χ, ψ are polynomials in t.

If x, y are ordinary rectangular point-coordinates, every straight line in the xy plane will be given by an equation of the form

$$y + ux + v = 0. \tag{2}$$

We may call u, v *the coordinates of the straight line*. Then $(-u)$ is the trigonometric tangent of the angle which the line makes with the x-axis, and $(-v)$ is the y-intercept (see Fig. 22). Let us think of points and lines as of equal importance; and let us give equal attention to point coordinates and line coordinates. This will be especially important later on. Then we may say that *the equation* $y + ux + v = 0$ *indicates the united position of the line* (u, v) *and of the point* (x, y), i. e., that the point lies on the given line, and the line goes through the given point.

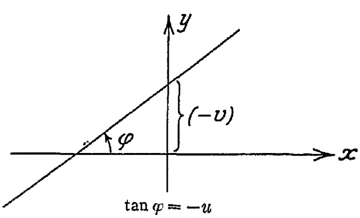


Fig. 22.

In order now to interpret the equation (1) geometrically, let us identify it with (2). This can be done in *two essentially different ways* which we shall consider, separably.

A. Let us consider the equations

$$(3\text{a}) \qquad y = \frac{\varphi(t)}{\psi(t)}, \qquad x = \frac{\chi(t)}{\psi(t)},$$

$$(3\text{b}) \qquad u = \lambda, \qquad v = \mu.$$

If t is variable, the equations (3a) represent, a *well determined rational curve of the $x\,y$ plane*, which is called the *normal curve of equation* (1). Since every point on it corresponds to a definite value of t, a certain *scale of* values of t is defined upon it. By means of (3a) we can calculate as many points as we please; and hence we can draw the normal curve, with its scale, as accurately as we please, say on millimeter paper. For every definite pair of values of λ and μ (3b) represents a *straight line* of the plane. From what has been said, it follow that (1) shows, that the point t of the normal curve lies upon this straight line. Thus *we obtain all the real roots of* (1) *if we find all the real intersections of the normal curve with this line and read off their parameter values on the curve scale.* The normal curve is determined, once for all, by the form of equation (1), regardless of the special values which the parameters λ, μ may have. For every equation with definite λ, μ there is, then, one straight line which represents it, in the manner described above, so that, in general, all the straight lines in the plane come into play, whereas before (pp. 87—88) only horizontal lines were used.

As an illustration, let us take the *quadratic equation*

$$t^2 + \lambda t + \mu = 0.$$

The normal curve here is given by the equations

$$y = t^2, \; x = t \qquad \text{or} \qquad y = x^2,$$

i.e., the normal curve is the parabola shown in Fig. 23, with the scale there indicated.

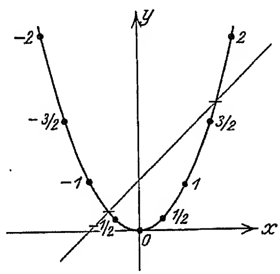


Fig. 23.

We can at once read off the real roots of our equation as the intersections with the line $y + \lambda x + \mu = 0$. In particular, the figure shows that the two roots of the equation $t^2 - t - 1 = 0$ lie between $\frac{3}{2}$ and 2, and between $-\frac{1}{2}$ and -1, respectively. The essential advantage of this method, over that given on pp. 87—88, is that we can now *solve all quadratic equations with one and the same parabola*, if we make use of all the straight lines in the plane. Thus, if we wish to solve, approximately, a considerable number of equations, one can apply this method very effectively.

In a similar way one can treat the totality of *cubic equations*, all of which can, by a linear transformation, be thrown into the *reduced form*

$$t^3 + \lambda t + \mu = 0.$$

The normal curve here is the cubical parabola

$$y = t^3, \quad x = t \qquad \text{or} \qquad y = x^3$$

sketched in Fig. 24. This method also seems to me to be usable in the schools. The pupils would certainly derive pleasure from drawing such curves.

B. The second method of interpreting (1) is got from the first by applying the *principle of duality*, i.e., by interchanging point and line coordinates. To that end, let us write the terms of (2) in reverse order:

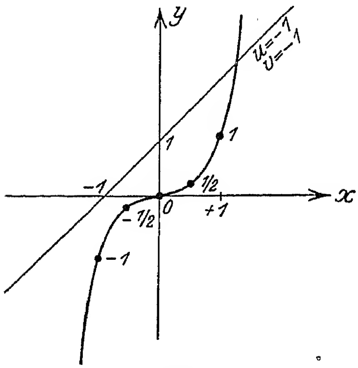


Fig. 24.

$$v + xu + y = 0$$

and identify it, in this form, with (1) by setting

(4a) $\quad v = \dfrac{\varphi(t)}{\psi(t)}, \qquad u = \dfrac{\chi(t)}{\psi(t)},$

(4b) $\quad x = \lambda, \qquad y = \mu.$

If t is variable, the equation (4a) represents a family of straight lines which will envelope a definite curve, the *normal curve of* (1), in the new interpretation. It is a *rational class curve*, since it is represented, in line coordinates, by rational functions of a parameter. Every tangent, and hence the corresponding point of tangency, is determined by a definite value of t, so that one gets again a *scale on the normal curve*. By drawing a sufficient number of tangents according to (4a), we may draw both curve and scale with any desired degree of exactness. Each parameter-pair λ, μ determines, by virtue of (4b), a point in the xy plane, through which, by virtue of (1), the tangent t of the normal curve (4a) must pass. *We obtain, therefore, all the real roots of* (1) *by reading off the parameter-values t belonging to all the tangents to the normal curve which go through the point $x = \lambda$, $y = \mu$.* As before, the normal curve is completely determined by the *form* of equation (1). Every equation of this form will be represented, for given values of the parameters λ, μ, by a certain point in the plane, or, if we wish, by its position with respect to the curve.

Let us illustrate by means of the same examples as before. Corresponding to the *quadratic* equation

$$t^2 + \lambda t + \mu = 0$$

the normal curve will be the envelope of the straight lines

$$v = t^2, \qquad u = t.$$

This envelope, again, is a *parabola* with its vertex at the origin. The graph, drawn on fine cross section paper exhibits immediately the real

roots of $t^2 + \lambda t + \mu = 0$ as parameters t of the tangents drawn to the parabola from the point λ, μ (see Fig. 25).

For the *cubic equation*

$$t^3 + \lambda t + \mu = 0$$

the normal curve

$$v = t^3, \qquad u = t$$

will be a curve of the *third class* with a cusp at the origin, shown in Fig. 26.

We can present this method somewhat differently. If we examine the so-called *trinomial equation*

$$t^m + \lambda t^n + \mu = 0,$$

we may represent the *system of tangents to the normal curve* by means of the parameter equation

$$f(t) = t^m + x t^n + y = 0.$$

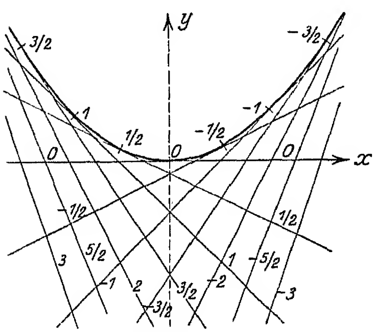


Fig. 25.

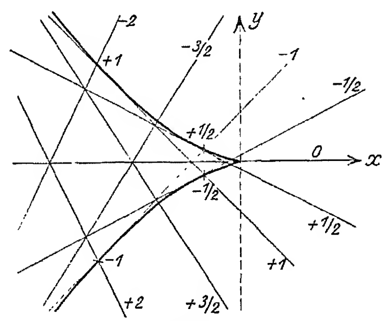


Fig. 26.

The equation of the normal curve in point coordinates may be found, as usual, by eliminating t between the last equation and the equation obtained by differentiation with respect to t:

$$f'(t) = m t^{m-1} + n x t^{n-1} = 0$$

for the normal curve, as the envelope of the system of straight lines, is the locus of the intersection of each of these lines with the neighboring line (for t and $t + dt$). If, instead of eliminating t, we express x and y as functions of t from these two equations, we find

$$\text{(5a)} \qquad x = -\frac{m}{n} t^{m-n}, \qquad y = \frac{m-n}{n} t^m,$$

which are the *point equations of the normal curve.*

As normal curves for the quadratic and the cubic equations which were selected above as examples, one finds in this way, respectively,

$$x = -2t, \qquad y = t^2$$

$$x = -3t^2, \qquad y = 2t^3.$$

These are the curves which are sketched in Figs. 25 and 26.

Let me emphasize the fact that this method is put to practical use by *C. Runge*, in his lectures and exercises, and that it has proved itself *especially appropriate for the actual solution of equations.* We might profitably use one or the other of these graphical methods in school instruction.

If we now compare with each other the two methods which we have developed, we find that, for at least one definite and very important purpose, the *second offers a distinct advantage*, namely, *when one seeks a visible representation of all the equations of a definite type which have a given number of real roots.* Such totalities are represented, according to the first method, by *systems of straight lines*; according to the second, however, by *fields of points*. But because of the peculiar nature of our geometric perception, or of our habit, the latter are essentially easier to grasp than are the former.

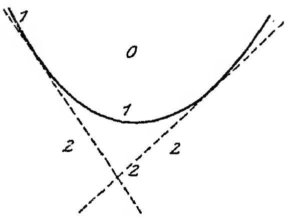


Fig. 27.

I shall show at once, by means of the *example of the quadratic equation*, what can be done in this direction (see Fig. 27). From all points outside of the parabola two tangents can be drawn to the curve; from points within, none. *Hence these two regions represent the manifolds of all equations with two roots and with no roots, respectively.* For all *points of the parabola itself* there is only a single tangent, which can be counted twice. *The normal curve itself is, then, in the general case, the locus of those points whose coordinates λ, μ yield equations with two equal roots,* so that we may call it the *discriminant curve.*

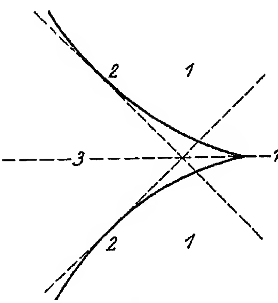


Fig. 28.

In the case of the *cubic equation,* we see that from a point inside the angle of the normal curve one can draw *three* tangents to the curve. This is obvious for points on the median line, because of symmetry; and the number cannot change when the point varies, provided it does not cross the curve. If the point (x, y) moves to the curve, two of the tangents coincide; if it moves into the region outside the curve, both of these tangents become imaginary and there remains but one real tangent. *Accordingly, the region inside the angle of the normal curve represents the totality of cubic equation with three different real roots; the region outside, equations with only one real root; while to the points on the curve itself correspond the equations with one simple and one double real root.* Finally, a triple tangent goes through the *cusp*, corresponding to the *single equation* $t^3 = 0$, *with a single triple root.* Figure 28 makes this obvious at a glance.

The pictures become much more interesting and show more, if, as is customary in algebra, we *impose definite restrictions upon the roots,* in particular, if we *inquire about all the real roots lying within a given interval* $t_1 \leqq t \leq t_2$. As you know, the general answer to this question is furnished by *Sturm's theorem*. We can, however, easily complete our drawing so that it will give a satisfying graphical solution of this general question also. For this purpose we simply *add to the normal curve the tangents to it determined by the parameter values* t_1, t_2 and consider the division of the plane into fields which these tangents bring about.

To carry through these considerations for the quadratic equation, we must *determine the number of tangents which touch the parabolic arc between* t_1 *and* t_2. Through every point of the triangle (see Fig. 29) bounded by the parabolic arc and these two tangents there are obviously two tangents. If the point crosses either of the tangents t_1, t_2, one of the tangents through it will touch the parabola beyond the arc (t_1, t_2), and so will be lost for our purpose. Tangents from points which lie within the two crescent shaped areas bounded by the parabola and the tangents t_1, t_2 touch the parabola outside the arc $(t_1 t_2)$; and from points within the parabola there are no real tangents at all. The two parabolic arcs $t \leqq t_1$ and $t \geqq t_2$ are thus of no significance in effecting the desired subdivision of the plane. There remain, then, only those lines in the figure which are drawn full; these, together with the numbers assigned to them, *give at a glance exact information as to the manifolds of quadratic equations which have* 2, 1, *or* 0 *real roots between* t_1 *and* t_2.

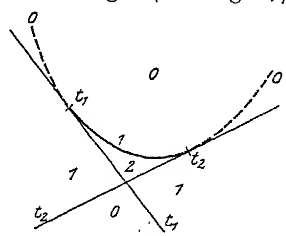


Fig. 29.

We may proceed similarly with the *cubic equation* (see Fig. 30). Let us take, say, $t_1 > 0$, $t_2 < 0$. Again we draw the tangents with these parameter values and examine the subdivisions of the plane brought about by them and the arc of the normal curve which lies between t_1 and t_2. Through every point in the four-cornered region at the cusp there will be three real tangents which touch between t_1 and t_2. If point crosses either of the tangents t_1, t_2, there is a loss of one tangent of this character. When it crosses the normal curve two are lost. From these considerations we obtain the picture, shown in Fig. 30, of the *regions of the plane which correspond to equations with three, two, one, or no roots lying between* t_1 *and* t_2. In order to see the great usefulness of the *graphical method*, one need only make a single attempt to picture

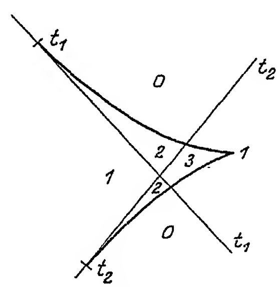


Fig. 30.

abstractly this classification of cubic equations, without making any appeal whatever to space perception; it will require a disproportionately great amount of time. And the proof, which here becomes evident by a glance at the picture, will not be at all easy.

Now as to the *relation of this geometric method to the well known algebraic criteria of Sturm, Cartesius, and Budan-Fourier* I remark, merely, that the geometric method includes them all, for equations of the types which we have considered. You will find these relations carried out more fully in my article[1] *"Geometrisches zur Abzählung der Wurzeln algebraischer Gleichungen"* and in *W. Dycks "Katalog mathematischer Modelle"*[2]. I am glad to take this occasion to refer you to this catalog. It was published on the occasion of the exposition, in Munich, in 1893, by the German Mathematical Society, and remains today the best means of orientation in the field of mathematical models.

3. Equations with three parameters λ, μ, ν

Finally, I shall also show you that one can apply analogous considerations to equations with three parameters. We shall need to use *space of three dimensions instead of the plane.* It will suffice if I consider the *special equation of four terms*

$$t^p + \lambda t^m + \mu t^n + \nu = 0\,.$$

The method of procedure can be applied immediately to equations of other forms.

In addition to this equation, we shall use the condition, from space geometry, that a point (x, y, z) and a plane with the plane coordinates (u, v, w) shall be "in united position", i. e., that the plane (u, v, w) shall contain the point (x, y, z). This condition is

$$z + ux + vy + w = 0 \tag{2}$$

or

$$w + xu + yv + z = 0\,. \tag{3}$$

We now identify this equation, written in the one form or the other, with (1) and we obtain, exactly as before, two mutually dual interpretations.

Let us then set

$$z = t^p, \qquad x = t^m, \qquad y = t^n. \tag{2a}$$

These equations determine a certain *space curve*, the *normal curve of the four-term equation* (1), *together with a scale of the values* t. Then we

---

[[1] Reprinted in Klein, F., Gesammelte Mathematische Abhandlungen, vol. II, pp. 198—208.]

[2] A catalogue of mathematical and mathematical-physical models, apparatus, and instruments (Munich, 1892), also a supplement to this (Munich, 1893).

consider the plane which is determined by the coefficients λ, μ, ν, of (1):

$$(2\text{b}) \qquad u = \lambda, \qquad v = \mu, \qquad w = \nu.$$

Then equation (1) says that the *real roots of the proposed equation are identical with the parameter values t of the real intersections of the normal curve* (2a) *with the plane* (2b).

If we choose the method dual to the preceding, we must put

$$(3\text{a}) \quad w = t^p, \quad u = t^m, \quad v = t^n.$$

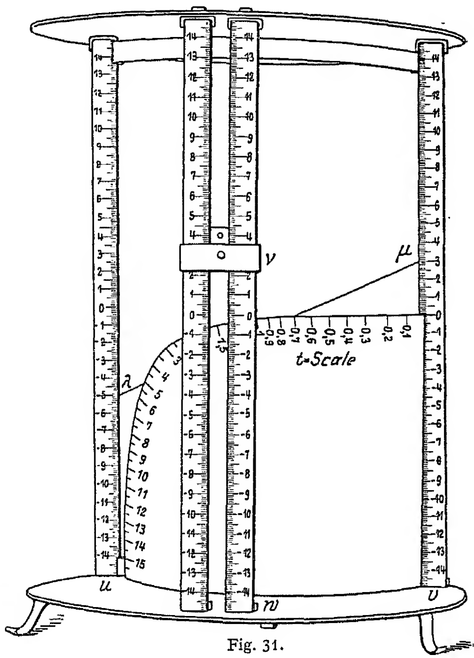


Fig. 31.

These equations represent, for variable t, a *simple infinity of planes*, which we can look upon as the *osculating planes of a definite space curve associated, as before, with a scale of parameter values t*. This will be a *normal "class curve"*, being expressed in plane coordinates, in distinction from the previous *normal* "order curve", which was given in point coordinates. If we now consider, in conjunction with the first curve, also the point

$$(3\text{b}) \quad x = \lambda, \quad y = \mu, \quad z = \nu,$$

it follows that *the real roots of* (1) *are identical with the parameter values t of those osculating planes of the normal class curve* (3a) *which pass through the point* (3b).

Let us next illustrate these two interpretations by *concrete examples*. We have, in our collection, models for both of them, which I shall now put before you.

The *first method* was used by *R. Mehmke*, in Stuttgart, in the *construction of an apparatus for the numerical solution of equations*. His model is a brass frame (see Fig. 31) in which you will notice three vertical rods carrying scales, and into which one can fit curved templates, or stencils, of the normal curves of equations of degree three, four, or five, (after these have been reduced to four terms). Note, however, that while our exposition presupposed the ordinary rectangular coordinate system, Mehmke has so *determined his coordinate system that the appropriate plane coordinates*, i.e., the coefficients u, v, w of the equation of the plane (2), *are precisely the intercepts which this plane makes on the*

scales of the three vertical rods and which one can read off there. In order, now, to make possible the fixation of a definite plane $u = \lambda$, $v = \mu$, $w = \nu$, a peep-hole is provided on the w rod, which one sets at the reading ν of that scale, while one joins by a stretched string the readings of the scales on the u and v rods, respectively. The rays joining the peep-hole with this string make our plane, and *by looking through the peep-hole one can observe directly the intersections of this plane with the normal curve as the apparent intersections of the string with the template.* Their parameter values, *the desired roots of the equation, are read at the same time on the scale of the normal curve, which is affixed to the template.* The practical usableness of this apparatus depends, of course, upon the carefulness of its mechanical construction, but the limited power of accommodation of the human eye would, at best, make it very doubtful.

For the *second method*, a model was prepared by *Hartenstein* in connection with his state examination. It has to do with the so-called *reduced form of the equation of degree four*, that is,

$$t^4 + \lambda t^2 + \mu t + \nu = 0,$$

to which every biquadratic equation can be reduced. I shall present this method in a form somewhat different from the one I used for the two-parameter equation (p. 91). In the present case we have to consider a simple infinity of planes whose plane coordinates are given in (3a) and whose point equations would be written as follows:

$$f(t) \equiv t^4 + xt^2 + yt + z = 0. \tag{4}$$

The envelope of these planes is the system of the straight lines in which each plane $f(t) = 0$ meets the neighboring plane $f(t + dt) = 0$, i.e., the *developable surface whose equation is obtained by eliminating t between $f(t) = 0$ and $f'(t) = 0$.* But in order to obtain the *normal curve* we must seek the *osculating configuration of the system of planes*, i.e., the *locus of the points of intersection of three successive planes.* This locus is, as you know, the *cuspidal edge of that developable surface and its coordinates are found, as functions of t, from the three equations $f(t) = 0$, $f'(t) = 0$, $f''(t) = 0$.* In our case these three equations are:

$$\begin{aligned} t^4 + xt^2 + yt + z &= 0 \\ 4t^3 + x \cdot 2t + y &= 0 \\ 12t^2 + x \cdot 2 &= 0, \end{aligned}$$

and one finds from them:

$$x = -6t^2, \qquad y = 8t^3, \qquad z = -3t^4. \tag{5}$$

These expressions represent the *point equation of the normal class curve of* (4) whose *plane equation*, by (3a), may be written in the form

$$w = t^4, \qquad u = t^2, \qquad v = t. \tag{6}$$

Both forms are of degree four in t. Hence *the normal curve is both of order four and of class four.*

In order to study it more in detail, let us consider *a few simple surfaces* which pass through it. In the first place, the expressions (5) satisfy identically in t the equation

$$z + \frac{x^2}{12} = 0,$$

Hence our normal curve lies upon a *parabolic cylinder of order two whose generators are parallel to the y-axis.* Likewise, we have the relation

$$\frac{y^2}{8} + \frac{x^3}{27} = 0,$$

so that this *cubic cylinder, whose generators are parallel to the z-axis,* also goes through our normal curve. *Moreover, the normal curve is the finite intersection of these two cylinders.* With these facts in mind, one can form an approximate picture of the course of the normal curve. *It is a skew curve, symmetric to the x z plane, having a cusp at the origin* (see Fig. 32).

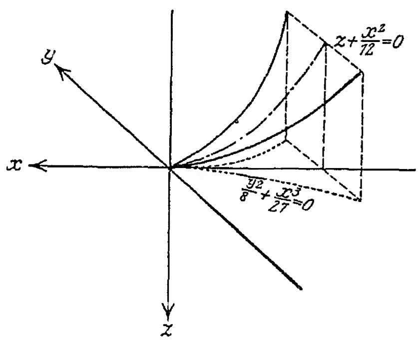


Fig. 32.

Again the *quadric surface*

$$\frac{x \cdot z}{6} - \frac{3y^2}{64} = 0$$

goes through our normal curve; for, by (5), this equation is also satisfied identically in t. From it, and the equation of the cubic cylinder, we find another linear combination which represents an especially important *surface of the third degree* passing through the normal curve:

$$\frac{xz}{6} - \frac{y^2}{16} - \frac{x^3}{216} = 0.$$

Let us now consider the *developable surface* whose cuspidal edge is the normal curve, and which we can define as the *totality of the tangents* to the normal curve. The tangent at the point t to any space curve

$$x = \varphi(t), \qquad y = \psi(t), \qquad z = \chi(t)$$

is given by the equations

$$x = \varphi(t) + \varrho\varphi'(t), \qquad y = \psi(t) + \varrho \cdot \psi'(t), \qquad z = \chi(t) + \varrho\chi'(t),$$

in which ϱ is a parameter. For the direction cosines of the tangents to the curve are to each other as the derivatives of the coordinates with respect to t. If t is thought of as variable, we have in these equations, with two parameters t, ϱ, the representation of the developable surface.

All this follows from well known theorems of space geometry. For our curve (5) we get, in particular, the following *equations* for the *developable surface:* If we call the coordinates of its points (X, Y, Z) to distinguish them from the coordinates of the curve, the equations of the developable are

$$(7)\qquad \begin{cases} X = -6(t^2 + 2\varrho t) \\ Y = 8(t^3 + 3\varrho t^2) \\ Z = -3(t^4 + 4\varrho t^3). \end{cases}$$

Now this surface is the basis of the Hartenstein model, its straight lines being represented by stretched threads (see Fig. 33).

The parameter representation offers the best starting point for the discussion and the actual construction of the surface. Indeed, it is only from force of habit that we inquire about the *equation of the surface* itself. We can obtain it by eliminating ϱ and t from (7). I shall give you the simplest procedure for this without giving the details of the inner meaning of the several steps. From (7) we form the combination

$$Z + \frac{X^2}{12} = 12\varrho^2 t^2,$$

$$\frac{X \cdot Z}{6} - \frac{Y^2}{16} - \frac{X^3}{216} = 8\varrho^3 t^3,$$

both of which vanish on the curve itself (for $\varrho = 0$). If we equate these to zero, we obtain two of the surfaces mentioned above which pass through the curve. Eliminating the product ϱt from these equations, we find the *equation of the developable surface*

$$\left(Z + \frac{X^2}{12}\right)^3 - 27\left(\frac{X \cdot Z}{6} - \frac{Y^2}{16} - \frac{X^3}{216}\right)^2 = 0.$$

The surface is thus of order six; but it is composed of the plane at infinity and a surface of order five.

As to the *meaning of this formula,* I make the following remark for those who are acquainted with the subject. The expressions in the two parentheses are the *invariants of the biquadratic equation*

$$t^4 + Xt^2 + Yt + Z = 0,$$

with which we started. These play an important role in the theory of elliptic functions and they are designated there, in general, by g_2 and g_3. The left side of the equation of our surface, $\Delta = g_2^3 - 27\, g_3^2$, is, as you know, the *discriminant of the biquadratic equation,* which indicates, by its vanishing, the presence of a repeated root. *Our developable surface is therefore the discriminant surface of the biquadratic equation,* i.e., the totality of the points for which it has a double root.

After these theoretical explanations, the construction of a thread model for our surface offers no essential difficulty. By means of the parameter equations (7) we may determine, say, the points in which

those tangents which we wish to represent intersect certain fixed planes. We then stretch threads between these planes, which are made out of wood or cardboard. But it requires long trial and great skill to make the model really beautiful and usable, and to bring out the entire interesting course of the surface and of its cuspidal edge, as in the model before us. The sketch on page 99 (see Fig. 33) shows the surface with its straight lines; AOB is the cuspidal edge [see the figure p. 97[1]].

You notice on the model a *double curve* (COD) *along which two sheets of the surface intersect.* This curve is simply the following parabola of the XZ plane:

$$Y = 0, \quad Z - \frac{X^2}{4} = 0.$$

Only one half (CO) of this parabola, namely that for $X < 0$, appears, however, as the *intersection of real sheets,* while *the other half lies isolated* in space. This phenomenon is by no means surprising to those who are accustomed to illustrate the theory of algebraic surfaces by concrete geometric representations. It is a common thing, there, for *real branches of double curves* to appear both as *intersections of real sheets* and also in part *isolated.* In the latter case we regard them as *real intersections of imaginary sheets* of the surface. The corresponding phenomenon in the plane is more generally known. In that case, in addition to the ordinary double points of algebraic curves, which appear as intersections of real branches of the curve, there are also the apparently isolated double points, which may be regarded as the intersections of imaginary branches.

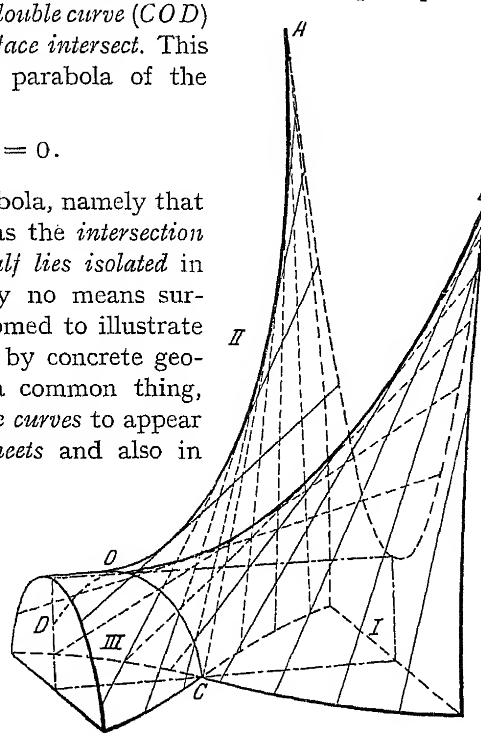


Fig. 33.

Let us now make clear in detail, what this surface with its cuspidal edge, the normal curve, can do for us. We think of the normal curve with its associated scale, or, better, we affix to each tangent its parameter value t, which also belongs to the point of tangency. *If, now, someone gives us a biquadratic equation with definite coefficients* (x, y, z), *we need only to pass through the corresponding point* (x, y, z) *the osculating plane to the normal curve, or, what would be the same thing, the tangent*

[1] The Hartenstein string model was put upon the market by the firm of M. Schilling in Leipzig. A dissertation by R. Hartenstein entitled: *Die Diskriminantenfläche der Gleichung vierten Grades* goes with the model Leipzig, Schilling, 1909.

plane to the discriminant surface, to obtain the real roots as the parameter values of the points of contact with the curve, or the parameter values of the corresponding tangents, as the case may be. Since the osculating plane cuts the curve where it touches it, every point of contact of an osculating plane with the curve is projected from the point (x, y, z) as an apparent point of inflexion on the curve, and conversely. *Consequently, the real roots of the biquadratic equation are, finally, the parameter values t of the apparent inflexion points of the normal curve, viewed from the point (x, y, z) in space.*

Now it is, of course, quite difficult for the unpractised eye to determine with certainty from the model either the planes of osculation or the apparent inflexions of the curve. But the model exhibits with immediate clearness the next important thing, *the classification of all biquadratic equations according to the number of their real roots.* Let us see, by an abstract examination of equations, just what cases one might expect. If $\alpha, \beta, \gamma, \delta$ are the four roots of the real biquadratic equation (4), then $\alpha + \beta + \gamma + \delta = 0$, because of the vanishing of the coefficient of t^3. So far as the reality of the roots is concerned, the following principal cases are possible:

I. *Four real roots.*

II. *Two real, and two conjugate complex roots.*

III. *No real, and two pairs of conjugate complex roots.*

If, now, two equations of the type I are proposed, with roots $\alpha, \beta, \gamma, \delta$ and $\alpha', \beta', \gamma', \delta'$, respectively, then one certainly could transform $\alpha, \beta, \gamma, \delta$ continuously into $\alpha', \beta', \gamma', \delta'$, respectively, through systems of values whose sum is always zero. At the same time, the one equation would transform continuously into the other, through equations always of the same type, i. e., all equations of type I make up a connected continuum, and the same is true for the other two types. *Our model must therefore exhibit space partitioned into three connected parts such that the points in each part correspond to equations of one type.*

Let us now consider the *transition cases between these three sorts.* Type I goes over into II through equations which have *two different real roots and one double (i. e. two coincident) real root,* which we shall indicate symbolically by $2 + (2)$; similarly we have between II and III the transition case of *one real double root and two complex roots,* which may be indicated by (2). *To both of these sorts there must correspond, in our model, regions of the discriminant surface,* which, indeed, pictures all equations with coincident roots. Considerations similar to those above would show *that to each type there must correspond a connected region of this surface.* Now, again, these two groups, $2 + (2)$ and (2), go over into each other by means of cases with *two real double roots,* symbolically: $(2) + (2)$; the points for which *two* pairs of roots move thus into coincidence must belong simultaneously to *two* sheets of the discriminant

surface, that is, to the *non isolated branch of the double curve. Accordingly, the discriminant surface falls into two parts, separated by a branch of the double curve; one of these parts,* 2 + (2), *separates the space regions I and II, the other,* (2), *the space regions II and III.* In order to see, now, how the normal curve lies, we notice that, because of its property as a cuspidal edge, *three tangent planes must merge into one* (*the osculating plane*) at each point on it, so that we have the case of a triple and a simple real root: 1 + (3). This can happen only when one of the simple roots becomes equal to the double root. Consequently, *the cuspidal edge must lie entirely on the first part,* 2 + (2), *of the surface.* In the *cusp* of the cuspidal edge ($x = y = z = 0$) we have a quadruple real root, which can arise from the case (2) + (2) through the coincidence of the two double roots. *In fact, the cusp, O, of the cuspidal edge lies also on the double curve.* Finally, *as to the isolated branch of the double curve,* it lies entirely in the space region III and is characterized by the fact that on it the *two pairs of conjugate complex roots merge into one complex double root.* Both double roots are, of course, conjugate to each other.

You can recognize on our model all of the possible cases enumerated above. In the sketch (Fig. 33, p. 99), the interior of the surface to the right of the double curve is region I, to the left, region III; the exterior is region II. You will be able easily to become fully oriented by means of the following tabulation, which exhibits the number and the multiplicity of the real roots which correspond to the points of the several space, surface, and curvilinear regions. In this scheme, the digits not in parentheses denote the number of simple real roots, the others, as before, denote the multiplicity of repeated roots:

	I.		II.		III.
Region:	4		2		0
Discrim. surface:		2 + (2)		(2)	
Normal curve:		1 + (3)			
Double curve			(2) + (2)		(2imag. double roots).
Cusp:			(4)		

II. Equations in the field of complex quantities

We shall now remove the restriction to real quantities and shall operate in the field of complex quantities. Of course, we shall endeavor again only to emphasize those things which are susceptible of geometric representation to an extent greater than one finds elsewhere. Let us begin at once with the most important theorem of algebra.

A. The fundamental theorem of algebra

This is, as you know, the theorem *that every algebraic equation of degree n in the field of complex numbers has, in general, n roots, or, more*

accurately, that every polynomial $f(z)$, of degree n, can be separated into n linear factors.

All proofs of this theorem make fundamental use of the *geometric interpretation* of the complex quantity $x + iy$ in the xy plane. I shall give you the *train of thought of Gauss' first proof* (*1799*), which can be presented quite graphically. To be sure, the original exposition of Gauss was somewhat different from mine.

Given the polynomial

$$f(z) = z^n + a_1 z^{n-1} + \ldots + a_n,$$

we may write

$$f(x + iy) = u(x, y) + i \cdot v(x, y),$$

where u, v are real polynomials in the two real variables $x . y$. *The leading thought of Gauss' proof lies now in considering the two curves*

$$u(x, x) = 0 \quad \text{and} \quad v(x, y) = 0$$

in the xy plane, and in showing that they must have one point, at least, in common. For this point one would then have $f(x + iy) = 0$, that is, *the existence of a first "root" of the equation $f = 0$ would be proved.* For this purpose, it turns out to be sufficient, to investigate the *behaviour of both curves at infinity*, i. e., at a distance from the origin which is arbitrarily great.

If r, the absolute value of z, is very large, we may neglect the lower powers of z in $f(z)$, in comparison with z^n. If we introduce polar coordinates r, φ into the xy plane, i. e., if we set

$$z = r(\cos\varphi + i \sin\varphi),$$

we have, by De Moivre's formula

$$z^n = r^n(\cos n\varphi + i \sin n\varphi).$$

This expression is approached asymptotically by $f(z)$, as z increases in absolute value. It follows at once that u and v approach, respectively, asymptotically the functions

$$r^n \cos n\varphi, \qquad r^n \sin n\varphi.$$

Consequently the ultimate course of the curves $u = 0$, $v = 0$, at infinity, respectively, will be given approximately by the equations

$$\cos n\varphi = 0, \qquad \sin n\varphi = 0.$$

Now the curve $\sin n\varphi = 0$ consists of the n straight lines which go through the origin and make with the x-axis the angles 0, π/n, $2\pi/n, \ldots, (n-1)\pi/n$, whereas $\cos n\varphi = 0$ consists of the n rays through the origin which bisect these angles (Fig. 34 is drawn for $n = 3$). In the central part of the figure, the true curves $u = 0$, $v = 0$ can, of course, be essentially different from these straight lines; but they must approach the straight lines asymptotically as the lines recede from the

origin. We *can indicate their course schematically by retaining the straight lines outside of a large circle and replacing them by anything we please, inside the circle* (see Fig. 35). But no matter what the behavior of the

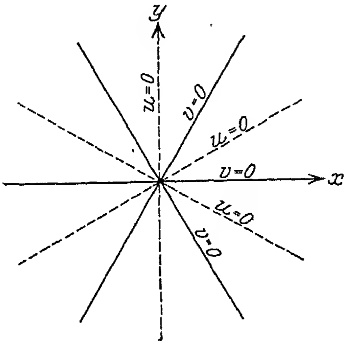


Fig. 34.

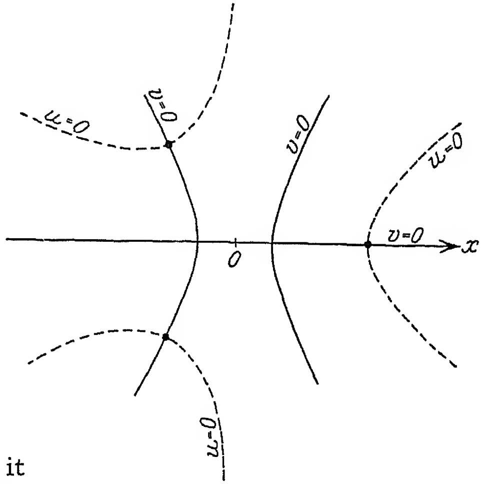


Fig. 35.

curves may be inside the circle, it is certain that, if one makes the circle about the origin sufficiently large, the branches u, v, outside the circle, must alternate, from which *it is graphically clear that these branches must cross one another inside the circle.* In fact, we can give a rigorous[1] proof of this assertion, — and this is the substance of Gauss' proof—if we use the continuity properties of the curves. The preceding argument, however, gives the essentials of the train of thought. If *one* such root has been found, we can divide out a linear factor, and we can then

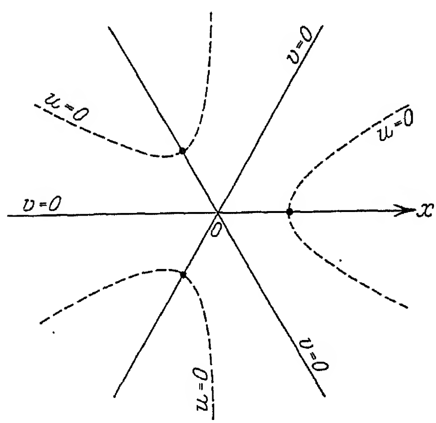


Fig. 36.

[1] It should be said here that Gauss does not dispense entirely with geometric considerations. The arithmetization of the proof which he contemplated in his dissertation was first given by A. Ostrowski (Göttinger Nachrichten, 1920, or vol. VIII of the materials for a scientific biography of Gauss, 1920). It is of historical interest that the first proof of the fundamental theorem was by D'Alembert. To be sure, there was an error in his proof, to which Gauss called attention. D'Alembert, namely, failed to distinguish between the upper limit of a function and its maximum, and he made use of the assumption, which in general is false, that a function of a complex variable actually attains its upper limit when this limit exists.

repeat the reasoning for the other polynomial factor of degree $(n-1)$. Continuing in this way, *we may finally break up $f(z)$ into n linear factors, i. e., we may prove the existence of n zeros.*

This method of reasoning will be much clearer if you *carry through the construction for special cases.* A simple example would be

$$f(z) = z^3 - 1 = 0.$$

In this case we obviously have

$$u = r^3 \cos 3\varphi - 1, \qquad v = r^3 \sin 3\varphi,$$

so that $v = 0$ consists simply of three straight lines, while $u = 0$ has three hyperbola-like branches. Figure 36 shows the three intersections of the two curves, which give the three roots of our equation. I recommend strongly that you work through other and more complicated examples.

These brief remarks about the fundamental theorem will suffice here, since I am not giving a course of lectures on algebra. Let me close by pointing out that the *significance of the admission of complex numbers into algebra* lies in the fact that it permits a general statement of the fundamental theorem. With the restriction to real quantities one can only say that the equation of degree n has n roots, or fewer, or perhaps none at all.

B. Equations with a complex parameter

The rest of the time which I have set aside for algebra I shall devote to the *discussion, by graphical methods, of all the roots (including the complex ones) of complex equations*, as was done earlier for the real roots of real equations. We shall limit ourselves, however, to equations with one complex parameter and we shall assume, furthermore, that this occurs *only linearly*. The study of a *simple conformal representation* will then give us all that is required.

Let $z = x + iy$ *be the unknown*, and $w = u + iv$ *the parameter*. Then the type of the equation to be considered has the form

$$\varphi(z) - w \cdot \psi(z) = 0 \tag{1}$$

where φ, ψ, are polynomials in z. Let n be the highest power of z that occurs. According to the fundamental theorem, this equation has for each definite value of w exactly n roots z which, in general, are different. Conversely, however, it follows from (1) that

$$w = \frac{\varphi(z)}{\psi(z)}, \tag{2}$$

i.e., *w is a single-valued rational function of z*, and it is said to be of *degree n*. If we should use, as geometric equivalent of equation (1),

simply the conformal representation which this function sets up between the z-plane and the w-plane, the many-valuedness of z as function of w would be visually disturbing. We may help ourselves here, as is always the case in function theory, by *thinking of the w-plane as consisting of n sheets, one over another, which are united in an appropriate manner, by means of branch cuts, into an n leaved Riemann surface.* Such surfaces are familiar to you all from the theory of algebraic functions. *Then our function establishes, between the points of the n-leaved Riemann's surface in the w-plane and the points of the simple z-plane, a one-to-one relation which is, in general, conformal.*

Before we begin a detailed study of this representation, it will be helpful if we set up certain conventions which will do away with the exceptional rôle played by infinite values of w and z, a role not justified by the nature of the case, and which will enable us to state theorems in general form. Inasmuch as these conventions are not so widely employed as they should be, you will permit me to say a word or two more about them than I otherwise should. We cannot be satisfied here when one speaks merely symbolically of an *infinitely distant point of the complex plane,* since such a conception gives no adequate concrete image, so that one must have recourse to special considerations or stipulations, in order to find out what corresponds, for an infinitely distant point, to a definite property of a finite point. *But we can secure all that is desired, if we replace the Gaussian plane, as picture of the complex numbers, once for all, by the Riemannian sphere.* For this purpose, we think simply of a sphere of diameter one, tangent to the xy plane, its south pole S being at the origin, and from its north pole N we project the plane stereographically upon the sphere (see Fig. 37). To every point $Q = (x, y)$ of the plane there corresponds uniquely the second intersection P of the ray NQ with the sphere; and, conversely, to every point P of the sphere, with the exception of N itself, there corresponds a unique point Q with definite coordinate (x, y). *Hence we can consider P as representing the number $x + iy$.* Now if P approaches the north pole N, in any manner, Q moves to infinity; conversely, if Q recedes to infinity in any manner, the corresponding point P approaches the single definite point N. *It seems natural, then, to look upon this point N,* which does not correspond to any finite complex number, *as the unique representative of all infinitely large $x + iy$, i.e., as the concrete picture of the infinitely*

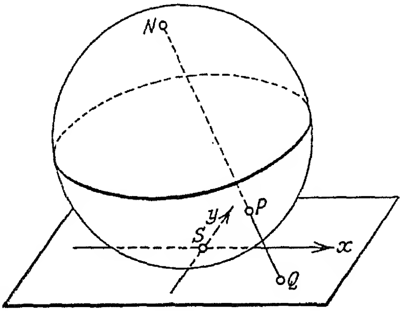


Fig. 37.

distant point of the plane, which is otherwise introduced only symbolically, and to affix to it outright the mark ∞. In this way we bring about, in the geometric picture, *complete equality between all finite points and the infinitely distant point.*

In order to return now to the geometric interpretation of the algebraic relation (1), we shall *replace the w plane also by a w-sphere.* Then our function will be represented by a *mapping of the z-sphere upon the w-sphere*, and, just as in the case of the mapping of the two planes, this is also *conformal*, since the stereographic mapping of the plane upon the sphere is, according to a well known theorem, conformal. To a single position on the w-sphere, there will then correspond, in general, n different positions on the z-sphere. In order to get a *one-to-one* relation we imagine, again, n sheets on the w-sphere, lying one above another, and united, in appropriate manner, by means of branch cuts, so as to form *an n-leaved Riemann surface over the w-sphere.* This picture presents no greater difficulty that that of the Riemann surface over the plane. *Thus, finally, the algebraic equation* (1) *is interpreted as a one-to-one relation, conformal in general, between the Riemann surface over the w-sphere and the simple surface of the z-sphere.* This interpretation obviously takes into account, also, infinite values of z and w which may correspond to each other or to finite values.

In order to make the greatest possible use this geometric device, we must take *a corresponding step in algebra*, one which shall do away with the exceptional role which infinity plays in the formulas, and this step is the *introduction of homogeneous coordinates.* We set, namely,

$$z = \frac{z_1}{z_2}$$

and consider z_1, z_2 as *two independent complex variables, both of which remain finite, and which cannot both vanish simultaneously.* Each definite value of z will then be given by infinitely many systems of values (cz_1, cz_2), where c is an arbitrary constant factor. We shall look upon all such systems of values (cz_1, cz_2) which differ only by such a factor, as the same *"position" in the field of the two homogeneous variables.* Conversely, for every such position there will be a definite value of z, with one exception: to the position (z_1 arbitrary, $z_2 = 0$) there will correspond no finite z; but if one approaches it from other positions, the corresponding z becomes infinite. *This one position is thus to be looked upon as the arithmetic equivalent of the one infinitely distant point of the z-plane or, as the case may be, of the z-sphere, and as carrying the mark* $z = \infty$.

In the same way, of course, we put also $w = w_1/w_2$. We shall now set up the *"homogeneous" equation between the "homogeneous" variables* z_1,

z_2 *and* w_1, w_2, which corresponds to equation (2). Multiplying by z_2^n in order to clear of fractions, we may write the equation in the form

$$(3) \qquad \frac{w_1}{w_2} = \frac{z_2^n \varphi\left(\frac{z_1}{z_2}\right)}{z_2^n \psi\left(\frac{z_1}{z_2}\right)} = \frac{\overline{\varphi}(z_1, z_2)}{\overline{\psi}(z_1, z_2)}.$$

In this equation, $\overline{\varphi}(z_1, z_2)$ and $\overline{\psi}(z_1, z_2)$ *are rational integral functions of* z_1 *and* z_2, since $\varphi(z)$ and $\psi(z)$ contain at most the nth power of $z = z_1/z_2$. Moreover they are *homogeneous polynomials (forms) of dimension* n. For each term z^i of $\varphi(z)$ or $\psi(z)$ is transformed into the term $z_2^n (z_1/z_2)^i = z_2^{n-i} z_1^i$, of dimension n, by clearing of fractions.

We come now to the detailed study of the functional dependence which our equation (1) *or, as the case may be,* (3) *establishes between* z *and* w. We shall apply consistently our two new aids, mapping upon the complex sphere and homogeneous coordinates. We shall have solved this problem when we can form a complete picture of the conformal relation between the z-sphere and the Riemann surface over the w-sphere.

First of all we must inquire as to the *nature and the position of the branch points of the Riemann surface.* I remind you here that a μ-fold branch point is one in which $\mu + 1$ leaves are connected. Since w is a single-valued function of z, we know the branch points when we know the points of the z sphere which correspond to them, which I am in the habit of calling the *critical or noteworthy points of the z-sphere.* To each of these there corresponds a certain *multiplicity* equal to that of the corresponding branch point. I shall now give, without detailed proof, the theorems which make possible the determination of these points. I assume that the rather simple functiontheoretic facts which enter into consideration here are in general familiar to you, though they may not be in the homogeneous form which I prefer to use. I shall illustrate in concrete graphical form the abstract considerations which I shall present to you, in this connection, by a series of examples.

A little calculation is necessary in order to obtain the *analogue, in homogeneous coordinates,* of the differential coefficient dw/dz. Differentiating equation (3) and omitting the bars over φ and ψ, we obtain

$$(3') \qquad \frac{w_2\, dw_1 - w_1\, dw_2}{w_2^2} = \frac{\psi\, d\varphi - \varphi\, d\psi}{\psi_2}.$$

We have also

$$d\varphi = \varphi_1 dz_1 + \varphi_2 dz_2,$$
$$d\psi = \psi_1 dz_1 + \psi_2 dz_2,$$

where

$$\varphi_1 = \frac{\partial \varphi(z_1, z_2)}{\partial z_1}, \qquad \varphi_2 = \frac{\partial \varphi(z_1, z_2)}{\partial z_2},$$
$$\psi_1 = \frac{\partial \psi(z_1, z_2)}{\partial z_1}, \qquad \psi_2 = \frac{\partial \psi(z_1, z_2)}{\partial z_2}.$$

On the other hand, from Euler's theorem for homogeneous functions of degree n, we have

$$\varphi_1 \cdot z_1 + \varphi_2 \cdot z_2 = n \cdot \varphi$$
$$\psi_1 \cdot z_1 + \psi_2 \cdot z_2 = n \cdot \psi;$$

consequently the numerator on the right side of (3′) may be written in the form

$$\psi d\varphi - \varphi d\psi = \begin{vmatrix} d\varphi, & d\psi \\ \varphi, & \psi \end{vmatrix} = \frac{1}{n} \begin{vmatrix} \varphi_1 dz_1 + \varphi_2 dz_2, & \psi_1 dz_1 + \psi_2 dz_2 \\ \varphi_1 z_1 + \varphi_2 z_2, & \psi_1 z_1 + \psi_2 z_2 \end{vmatrix}.$$

This expression, by the multiplication theorem for determinants, becomes

$$= \frac{1}{n} \begin{vmatrix} \varphi_1, & \varphi_2 \\ \psi_1, & \psi_2 \end{vmatrix} \cdot \begin{vmatrix} dz_1, & dz_2 \\ z_1, & z_2 \end{vmatrix}.$$

Thus (3′) goes over into the equation

$$\frac{w_2 dw_1 - w_1 dw_2}{w_2^2} = \frac{z_2 dz_1 - z_1 dz_2}{n \cdot \psi^2} (\varphi_1 \psi_2 - \psi_1 \varphi_2).$$

This constitutes the *basal formula of the homogeneous theory of our equation*, and the *functional determinant* $\varphi_1 \psi_2 - \varphi_2 \psi_1$, *of the forms* φ, ψ appears as a crucial expression for all that follows. Except for it and for the factor $z_2^2/(n\psi^2)$, one has on the right the differential of $z = z_1/z_2$, on the left that of $w = w_1/w_2$. Since for finite z and w the critical points are given by $dw/dz = 0$, as is well known, the following theorem appears *plausible*, but I shall here omit the proof. *Each μ-fold zero of the functional determinant is a critical point of multiplicity μ, i.e., there corresponds to it a μ-fold branch point of the Riemann surface over the w-sphere.* The chief advantage of this rule, as compared with those which are otherwise given, lies in the fact that it contains in one statement both finite and infinite values of z and w. It enables us also to make a precise statement concerning the *number of remarkable points*. The four derivatives, namely, are forms of dimension $n - 1$, and the functional determinant is therefore a form of dimension $2n - 2$. Such a polynomial always has $2n - 2$ zeros, if one takes into account their multiplicity. *Thus, if $\alpha_1, \alpha_2, \ldots, \alpha_\nu$ are the remarkable points of the z-sphere (i.e., if $\varphi_1 \psi_2 - \varphi_2 \psi_1 = 0$ for $z_1 : z_2 = \alpha_1, \alpha_2, \ldots, \alpha_\nu$) and if $\mu_1, \mu_2, \ldots, \mu_\nu$ are their respective multiplicities, then their sum is*

$$\mu_1 + \mu_2 + \cdots + \mu_\nu = 2n - 2.$$

By virtue of the conformal mapping, to these points there correspond the ν *branch points*

$$a_1, a_2, \ldots, a_\nu$$

on the Riemann surface over the w-sphere, which must necessarily lie separated on the surface, and about which $\mu_1 + 1$, $\mu_2 + 1$, $\ldots$, $\mu_\nu + 1$ leaves, respectively, must be cyclically connected. It should be noted,

however, that different ones of these branch points may lie over the same position on the w sphere, since $w = \varphi(z)/\psi(z)$ for $z = \alpha_1, \alpha_2, \ldots, \alpha_\nu$ may give the same value for w more than once. Over such a point, there would be two or more separate series of leaves, each series being in itself connected. Every such position on the w sphere is called a branch position; we shall denote them, in order, by $A, B, C, \ldots$. It should be noted that their number can be smaller than ν.

The statements thus far made furnish only a hazy picture of the *Riemann surface*. We shall now *build it up so that it can be more readily visualized. For this purpose, let us draw on the w sphere through the branch positions $A, B, C, \ldots$ an arbitrary closed curve $\mathfrak{C}$ without double points and of the simplest possible form* (see Fig. 38), *and distinguish the two spherical caps thus formed as the upper cap and the lower cap*. In all of the examples which we shall discuss later the points $A, B, C, \ldots$ will all be real and we shall then naturally select *as the curve $\mathfrak{C}$ the meridian great circle of real numbers*, so that each of our two partial regions will be a hemisphere.

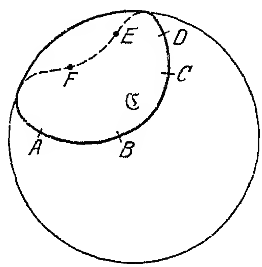


Fig. 38.

Returning to the general case we see that each pair of leaves of the Riemann surface which are connected, intersect along a *branch cut* which joins two branch points. As you know, the Riemann surface remains unchanged in essence if we move these cuts, leaving the end points fixed, that is, if we think of the same leaves as being connected along other curves, provided these join the same branch points. It is in just this variability that the great generality and also the great difficulty of the idea of the Riemann surface lies. In order to give the surface a definite form, which shall be susceptible of concrete visualization, *we move all the branch cuts so that all of them lie over the curve $\mathfrak{C}$, which passes through all the branch points*. It may be that several branch cuts lie over the same part of $\mathfrak{C}$, and none at all over other parts.

Now let us cut this entire complex of leaves, i. e., each individual leaf, along the curve $\mathfrak{C}$. Since we had already moved all the branch cuts into position over $\mathfrak{C}$, the incision just made passes along all of them, so that *our Riemann surface separates into $2n$ "half-leaves" entirely free from branches, n of them over each of the two spherical caps*. If we think of the half-leaves corresponding to the upper cap as being shaded, and those corresponding to the lower as not shaded, we can distinguish briefly, *n shaded and n unshaded half-leaves*. We can now describe *the original Riemann surface as follows*. On it *each shaded half-leaf meets only unshaded half-leaves, those with which it is connected along segments of the curve $\mathfrak{C}$ lying over AB, BC, ...; and, similarly, each unshaded half-leaf*

is connected along such segments of $\mathfrak{C}$ only to shaded half-leaves. However, more than two half-leaves may meet only at a branch point; and in fact around any μ-fold branch point, $\mu + 1$ shaded half-leaves would alternate with $\mu + 1$ unshades ones.

Since the mapping by means of our function $w(z)$ of the z sphere upon the Riemann surface over the w sphere is a one-to-one correspondence, we can immediately transfer to the z sphere the above conditions of connectivity. Because of continuity, the $2n$ half-leaves of the Riemann surface must correspond to $2n$ connected z regions, which we may call the shaded and the unshaded half-regions. These will be separated from one another by the n images of each of the segments $AB, BC, \ldots$ of the curve $\mathfrak{C}$ which the n-valued function $z(w)$ represents upon the z sphere. *Each shaded half-region meets only shaded half-regions along these image-curves, and each unshaded half-region meets only shaded ones. It is only in a μ-fold critical point that more than two half-regions can meet. At such a point $\mu + 1$ shaded and $\mu + 1$ unshaded half-regions come together.*

This division of the z sphere into partial regions will help us to follow in detail the course of the function $z(w)$ for a few simple characteristic examples. I shall begin with the simplest one possible.

1. The "pure" equation

We shall call the well known equation

$$z^n = w \tag{1}$$

a pure equation. Its solution is given formally by introducing the radical $z = \sqrt[n]{w}$. This gives us no information, however, regarding the functional relation between z and w. We shall proceed according to the general plan by introducing the homogeneous variables

$$\frac{w_1}{w_2} = \frac{z_1^n}{z_2^n},$$

and we shall consider the functional determinant of the numerator and denominator of the right side

$$\begin{vmatrix} n z_1^{n-1}, & 0 \\ 0, & n z_2^{n-1} \end{vmatrix} = n^2 z_1^{n-1} \cdot z_2^{n-1}.$$

This expression obviously has the $(n-1)$ *fold zeros* $z_1 = 0$ and $z_2 = 0$, or (in non-homogeneous form) $z = 0$ and $z = \infty$. These are the only critical points and they are of total multiplicity $2n - 2$. By our general theorem, therefore, *the only branch points of the Riemann surface over the w sphere* are at the positions $w = 0$ and $w = \infty$. By the equation $w = z^n$ these correspond to the two points $z = 0$ and $z = \infty$. Each of these two points has the *multiplicity* $n - 1$, so that n leaves are

cyclically connected at each of them. Let us now mark on the w sphere *the meridian of real numbers as the curve* $\mathfrak{C}$ and let us cut all the leaves of the Riemann surface along this meridian, after having appropriately displaced all of the branch cuts. Of the $2n$ hemispheres into which the surface separates we think of those over the *rear* half of the w sphere, that is, those which correspond to w values with *positive imaginary parts*, as shaded. Upon the meridian itself, we shall distinguish between the *half meridian of positive real numbers* (drawn full in Fig. 39) and that of the *negative real numbers* (*dotted*).

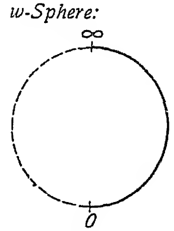


Fig. 39.

Now we must examine the *mappings of this meridian* $\mathfrak{C}$ *curve upon the z sphere*, where they bring about the characteristic division into half-regions. Upon the positive half meridian $w = r$, where r ranges through positive real values from 0 to ∞; for these values we have by a well known formula of complex numbers,

$$z = \sqrt[n]{w} = \left|\sqrt[n]{r}\right| \left(\cos\frac{2k\pi}{n} + i\sin\frac{2k\pi}{n}\right), \quad \text{where} \quad k = 0, 1, \ldots, n-1.$$

For the different values of k, this expression gives those n *half-meridians of the z sphere which make with the half-meridian of positive real numbers the angles* $0, 2\pi/n, 4\pi/n, \ldots, 2(n-1)\pi/n$. Thus these curves corres-

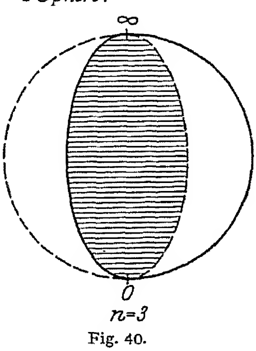


Fig. 40.

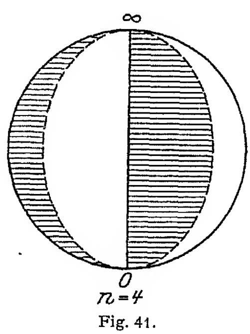


Fig. 41.

pond to the full drawn half of $\mathfrak{C}$. On the negative half-meridian of the w sphere we must set $w = -r = r \cdot e^{i\pi}$, where again $0 \leqq r \leqq \infty$. This gives

$$z = \sqrt[n]{w} = \left|\sqrt[n]{r}\right| \left(\cos\frac{(2k+1)\pi}{n} + i\sin\frac{(2k+1)\pi}{n}\right), \text{where } k = 0, 1, \ldots, n-1.$$

Corresponding to this we have, on the z sphere, those n *half-meridians which have the "longitude"* $\pi/n, 3\pi/n, \ldots, 2(n-1)\pi/n$, *which thus bisect the angles between the others*. Accordingly, the z sphere is divided into $2n$ congruent *sectors reaching from the north pole to the south pole*, similar

to the natural divisions of an orange. This division is exactly in accord with the general theory. In particular, it is only at the remarkable points, the two poles, that more than two half-regions meet. At each of these points $2n$ half-regions meet, corresponding to the multiplicity $n-1$.

As for the *shading of the regions*, we need to fix it for *one* region only. The remainder are then alternately shaded and unshaded. Now note that when we look at the shaded half of the w sphere (the rear) from the point $w=0$, the full drawn part of the boundary lies to the left, the dotted part to the right. Since we are concerned with a conformal mapping in which angles are not reversed, *each shaded portion of the z sphere, looked at from the correponding point $z=0$, must have the same property as to position, that is, it must have a full drawn boundary to the left, and a dotted one to the right.* With this we control completely the division of the z sphere into regions. Moreover, one notices a characteristic difference in the distribution of the regions upon two z hemispheres, according as n is even or odd, as can be clearly seen in Figs. 40 and 41 on p. 111 for the first cases $n=3$, $n=4$. Let me emphasize how necessary it was to go over to the complex *sphere* in order to get a full understanding of the situation. In the complex z *plane*, one would have a division into angular sectors by straight lines radiating from $z=0$, and it would not be at all so obvious that $z=\infty$ and $w=\infty$ have equal significance with $z=0$ and $w=0$, as critical point and branch point, respectively.

This furnished us with the essentials for exact knowledge of the functional relation between z and w. We need now study only the *conformal mapping of each of the $2n$ spherical sectors upon one or the other of the two w hemispheres.* But I shall not go into the details here. This case, as one of the simplest and most obvious illustrations, will be familiar ground to any one who has had to do with conformal representation. We shall see later (see p. 131) how to deduce from this methods for the numerical calculation of z.

Let us, however, settle here the important question as to the mutual relation *among the various congruent regions of the z sphere.* Speaking more exactly, $w=z^n$ takes on the same value at a point in each one of the n shaded regions. Can the corresponding values of z be expressed in terms of one another? We notice, in fact, that for $z'=z\cdot\varepsilon$ (where ε is any one of the nth roots of unity) $z'^n=z^n$, that is $w=z^n$ *takes the same value at all the n positions*

(2) $$z'=\varepsilon^\nu\cdot z=e^{\frac{2\nu i\pi}{n}}\cdot z\quad(\nu=0,1,2,\ldots,n-1).$$

These n values of z' must therefore be distributed so that just one of them lies in each of the n shaded regions of the z sphere, if z is taken

in *one* of the shaded regions and each of them must traverse one of these regions as z traverses its region. The same thing is true of the unshaded regions. Each of the substitutions (2) is represented geometrically by *a rotation of the z sphere through an angle $\nu \cdot 2\pi/n$ about the vertical axis* $0, \infty$, since, as is well known, multiplication in the complex plane by $e^{2\nu i\pi/n}$ denotes a rotation through that angle about the origin. Thus *corresponding points of our spherical regions, as well as the regions themselves, go over into one another by means of these n rotations about the vertical axis.*

If, then, we had determined at the start only one shaded partial region of the sphere, this remark would have furnished *all the similar partial regions*. In this we have made use only of the property of the substitutions (2) *that they transform equation* (1) *into itself* (*i.e.*, $z^n = w$ *into* $z'^n = w$) *and that their number is equal to the degree*. In the examples that follow, we shall always be able to give such linear substitutions at the outset, and by means of them to simplify the determination of the division into subregions.

By using the present example I should like to illustrate *an important general notion*, namely, the *notion of irreducibility for equations which contain a parameter w rationally*. We have already discussed irreducibility of equations with rational *numerical* coefficients in connection with the construction of the regular heptagon (p. 51 et seq.). *An equation* $f(z, w) = 0$ (e.g., our equation $z^n - w = 0$), *where $f(z, w)$ is a polynomial in z, whose coefficients are rational functions of w, is called reducible with respect to the parameter w, when f can be split into the product of two polynomials of the same sort, in each of which z really appears*

$$f(z, w) = f_1(z, w) \cdot f_2(z, w);$$

otherwise the equation is called irreducible with respect to w. The entire generalization, in comparison with the earlier conception, lies in the fact that the "*domain of rationality*" in which we operate and in which the coefficients of the admissible polynomials are to lie, consists of the *totality of rational functions of the parameter w instead of the totality of rational numbers*, in other words, that we pass from a numbertheoretic to a functiontheoretic conception.

If we illustrate this, for each equation $f(z, w) = 0$, by means of its Riemann surface, we can set up *a simple criterion for reducibility in this new sense.* If the equation, namely, is reducible, every system of the values z, w which satisfies it satisfies either $f_1(z, w) = 0$ or $f_2(z, w) = 0$; now the solutions of $f_1 = 0$ and $f_2 = 0$ are represented by means of their Riemann surfaces, which have nothing to do with each other, and, in particular, are not connected. Thus, *the Riemann surface which belongs to a reducible equation $f(z, w) = 0$ must break down into at least two separates pieces.*

According to this, we can now assert that the equation $z^n - w = 0$ *is certainly irreducible in the function theoretic sense.* For, on its Riemann surface, which we known exactly, all the n leaves are cyclically connected at each of its branch points. Moreover, the entire surface is mapped upon the unpartitioned z sphere. Hence such a breaking down cannot occur.

In connection with this, we can answer one of the popular problems of mathematics which we touched earlier (p. 51), *namely, that of the possibility of dividing an arbitrary angle φ into n equal parts, in particular, for $n = 3$, the possibility of trisecting an angle.* The problem is to give an exact construction *with ruler and compasses* for dividing into three equal parts *any angle φ whatever.* (It is easy, of course, to give a construction for a series of *special values of φ*). I shall give you the train of thought for the proof of the impossibility of trisecting an angle in the sense just mentioned, and I shall ask you to recall, in this connection, the proof of the impossibility of constructing the regular heptagon with ruler and compasses (see p. 51 et seq.). Just as at that time, we shall reduce the problem to that of the solution of an *irreducible cubic equation*, and we shall then show that this equation cannot be solved by a series of square roots; except that, now, the equation will contain *a parameter* (the angle φ), whereas, before, the coefficients were integers. Accordingly, *functiontheoretic irreducibility must replace numbertheoretic* irreducibility.

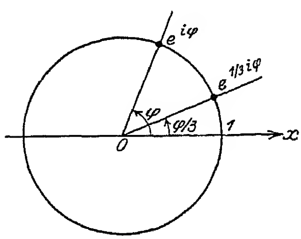


Fig. 42.

In order to set up the equation of the problem let us think of the angle φ as laid off from the positive real half-axis in the w plane (see Fig. 42). Then its free arm will cut the unit circle in the point

$$w = e^{i\varphi} = \cos\varphi + i\sin\varphi.$$

Our problem consists in finding, independently of special values of the parameter φ, a construction, involving a finite number of applications of the ruler and compasses, which shall give the point of intersection with the unit circle of the arm of the angle $\varphi/3$, i. e., the point

$$z = e^{\frac{i\varphi}{3}} = \cos\frac{\varphi}{3} + i\sin\frac{\varphi}{3}.$$

This value of z satisfies the equation:

$$(3) \qquad z^3 = \cos\varphi + i\sin\varphi,$$

and the analytic equivalent of our geometric problem consists in *solving this equation* (see p. 51) *by means of a finite number of square roots, one over another, of rational functions of* $\sin\varphi$ *and* $\cos\varphi$, since these quantities are the coordinates of the point w with which we start the construction

We must show, first, *that the equation* (3) *is irreducible in the function theoretic sense*. To be sure, this equation does not have just the form we assumed while explaining the notion, since, instead of the a *complex* parameter w that enters rationally, we have now two functions cos and sin of a *real* parameter φ, both of which appear rationally. As a natural extension here of our notion, we shall *call the polynomial* $z^3 - (\cos\varphi + i\sin\varphi)$ *reducible if it can be split into polynomials whose coefficients are likewise rational finctions of* $\cos\varphi$ *and* $\sin\varphi$; and we can, as before, assign a criterion for this. If we let φ assume all real values in (3), $w = e^{i\varphi} = \cos\varphi + i\sin\varphi$ will describe the unit circle of the w plane, to which the equation of the w sphere corresponds by stereographic projection. The curve which lies over this, on the Riemann surface of the equation $z^3 = w$, and which describes, in one stroke, all three leaves, is mapped by equation (3) uniquely upon the unit circle of the z sphere. Hence it can be regarded, in a sense as its *"one dimensional Riemann image"*. In the same way, we can obviously assign such a Riemann image to every equation of the form $f(z, \cos\varphi, \sin\varphi) = 0$ by taking as many copies of the unit circle with arc length φ as the equation has roots, and joining them according to the connectivity of the roots. It follows, just as before, *that the equation* (3) *can be reducible only when its one-dimensional Riemann image breaks down into separate parts*, and this is obviously not the case. *This proves the function theoretic irreducibility of our equation* (3).

Now, however, the former proof of the theorem, that a cubic equation with rational numerical coefficients is reducible if it can be solved by a series of square roots, can be applied literally to the present case of the function-theoretically irreducible equation (3) (see p. 51 et seq.). We need only to replace "rational numbers" there by "rational functions of $\cos\varphi$ and $\sin\varphi$". *This proves our assertion that the trisection of an arbitrary angle cannot be accomplished by a finite number of applications of a ruler and compasses.* Hence the endeavors of angle-trisection zealots must always be fruitless!

I pass on now to the treatment of a somewhat more complicated example.

2. The dihedral equation

The equation

$$(1) \qquad w = \frac{1}{2}\left(z^n + \frac{1}{z^n}\right).$$

is called the dihedral equation, for reasons that will appear later. Clearing of fractions, we see that its degree is $2n$. Introducing homogeneous variables we get

$$\frac{w_1}{w_2} = \frac{z_1^{2n} + z_2^{2n}}{2z_1^n \cdot z_2^n},$$

in which, in fact, forms of dimension $2n$ appear in numerator and denominator. The functional determinant of these forms is

$$\begin{vmatrix} 2n z_1^{2n-1}, & 2n z_2^{2n-1} \\ 2n z_1^{n-1} z_2^n, & 2n z_1^n z_2^{n-1} \end{vmatrix} = 4n^2 z_1^{n-1} z_2^{n-1} (z_1^{2n} - z_2^{2n}).$$

It has an $(n-1)$-fold zero at $z_1 = 0$ and at $z_2 = 0$; the other $2n$ zeros are given by

$$z_1^{2n} - z_2^{2n} = 0 \qquad \text{or:} \qquad \left(\frac{z_1}{z_2}\right)^n = \pm 1 .$$

If in addition to the n-th root of unity

$$\varepsilon = e^{\frac{2i\pi}{n}}$$

which we have already used, we introduce also the primitive n-th root of -1:

$$\varepsilon' = e^{\frac{i\pi}{n}},$$

the last $2n$ zeros are given by the equations

$$\frac{z_1}{z_2} = \varepsilon^\nu \qquad \text{and} \qquad \frac{z_1}{z_2} = \varepsilon' \cdot \varepsilon^\nu, \qquad (\nu = 0, 1, \ldots, n-1) .$$

Since the values of $z = z_1/z_2$ corresponding to them all have the absolute value one, they all lie therefore on the equator of the z sphere (corresponding to the unit circle of the z plane), at equal angular spacings of π/n. We have therefore *as critical points on the z sphere*:

(a) *the south pole $z = 0$ and the north pole $z = \infty$, each of multiplicity $n - 1$*;

(b) *the $2n$ equatorial points $z = \varepsilon^\nu$, $\varepsilon' \cdot \varepsilon^\nu$, each of multiplicity one.*

The sum of all the multiplicities is $2 \cdot (n-1) + 2n \cdot 1 = 4n - 2$, as is demanded by the general theorem on p. 108 for the degree $2n$. By virtue of equation (1) there will correspond to the remarkable points $z = 0$, $z = \infty$ of the z sphere, the position $w = \infty$ on the w sphere. Moreover, to all the points $z = \varepsilon^\nu$, corresponds the position $w = +1$; and, to all the points $z = \varepsilon' \cdot \varepsilon^\nu$ the position $w = -1$. There are, accordingly, *only three branch points* ∞, $+1$, -1 *on the w sphere.* These will lie as follows:

$w = \infty$ *two branch points of multiplicity* $n - 1$;
$w = +1$ n *branch points of multiplicity* 1;
$w = -1$ n *branch points of multiplicity* 1.

The $2n$ leaves of the Riemann surface group themselves therefore over the point $w = \infty$ in two separate series, each of n cyclically connected leaves; over $w = +1$ and $w = -1$ in n series, each of two leaves. The disposition of the leaves will become clear when we study the *corresponding subdivision of the z sphere* into half-regions.

To this end it will be well, as we remarked above, *to know the linear substitutions which transform equation* (1) *into itself*. As in the case of the pure equation, it is unchanged by the n substitutions

(2a) $$z' = \varepsilon^{\nu} \cdot z \, (\nu = 0, 1, \ldots \, n-1), \quad \text{where} \quad \varepsilon = e^{\frac{2i\pi}{n}},$$

since for these $z'^n = z^n$. Likewise, however, it is unchanged by the n additional substitutions

(2b $$z' = \frac{\varepsilon^{\nu}}{z} \, (\nu = 0, 1, \ldots \, n-1).$$

since these only change z^n into $1/z^n$.

We have therefore $2n$ linear substitutions of equation (1) into itself, exactly *as many as its degree indicates*. Thus, if we know for a given value w_0 of w *one* root z_0 of the equation, we know immediately $2n$ roots

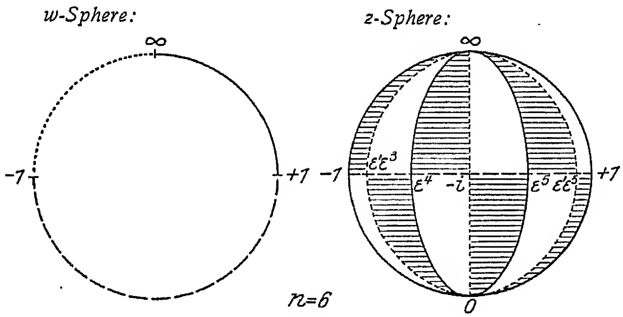


Fig. 43.

$\varepsilon^{\nu} \cdot z_0$ and ε^{ν}/z_0 $(\nu = 0, 1, 2, \ldots, n-1)$, in general all different, for which w has the same value w_0, i. e., we know *all* the roots of the equation when we have obtained the n-th root of unity ε.

Let us now proceed to examine the subdivision of the z sphere corresponding to cuts along the real meridian of the Riemann surface over the w sphere. In this, as in the previous example, we distinguish on the real meridian of the w sphere the three segments made by the branch points —that from $+1$ to ∞ (drawn full), that from ∞ to -1 (short dotted), and that from -1 to $+1$ (long dotted) (see Fig. 43). *To each of these three segments there correspond on the z sphere $2n$ different curvilinear segments which can be derived from any one of them by means of the $2n$ linear substitutions* (2). It will always suffice, therefore, to find one of them. Moreover all these segments must connect the critical points $z = 0, \infty, \varepsilon^{\nu}, \varepsilon' \cdot \varepsilon^{\nu}$, which we therefore mark on the z sphere. Just as in the previous case, their form is of a somewhat different type according as n is even or odd. It will suffice if we exhibit a definite case, say for $n = 6$. Fig. 43 shows the front half of the z sphere in orthogonal projection. One sees, on the equator, from left to right with spacings of

60°, $\varepsilon^3 = -1$, $\varepsilon^4, \varepsilon^5, \varepsilon^6 = 1$; and lying midway between the others, $\varepsilon' \cdot \varepsilon^3$, $\varepsilon' \cdot \varepsilon^4 = -i$, and $\varepsilon' \cdot \varepsilon^5$.

Now we shall see that the quadrant $+1 < z < \infty$ of the meridian of real z corresponds to the part of the real w meridian $+1 < w < \infty$ (full drawn). In fact, if we put $z = r$ and let r range through real values from 1 to ∞, then $w = \frac{1}{2}(z^n + 1/z^n) = \frac{1}{2}(r^n + 1/r^n)$ will vary also through real values that are always increasing, from 1 to ∞. We obtain n other full drawn curves on the z sphere, from this one, by means of the n linear substitutions (2a). But, as we saw in the previous example, these substitutions mean rotations of the sphere about the vertical axis $(0, \infty)$ through the angles $2\pi/n, 4\pi/n, \ldots, 2(n-1)\pi/n$. *We get in this way the n quarter-meridians from the north pole ∞ to the points ε^ν on the equator.* We get an additional full drawn curve if we apply the substitution $z' = 1/z$, which transforms the meridian quadrant from $+1$ to ∞ into the lower real meridianquadrant from $+1$ to 0. If we subject this quadrant to the n rotations (2a), the composition of which with $z' = 1/z$ gives the n substitutions (2b), we obtain, in addition, *the n meridian quadrants which join the south pole with the equatorial points ε^ν.* We have now in fact the $2n$ full drawn curves which correspond to the full drawn w meridian quadrant. In particular, for $n = 6$, they make up the three entire meridians into which the real meridian is transformed by rotations of $0^\circ, 60^\circ, 120^\circ$.

It is now also obvious that the totality of the values $z = \varepsilon' \cdot r$, where r again ranges through real values from $+1$ to ∞, corresponds to the dotted part of the real w meridian; for the equation (1) yields then:

$$w = \frac{1}{2}\left(\varepsilon'^n r^n + \frac{1}{\varepsilon'^n r^n}\right) = -\frac{1}{2}\left(r^n + \frac{1}{r^n}\right),$$

and this expression actually decreases through real values from -1 to $-\infty$. But $z = \varepsilon' \cdot r$ represents *the meridian quadrant from ∞ to the equatorial point ε^ν.* If we now apply to it the substitutions (2a), (2b), we find, as before, *that to the dotted part of the real w meridian there correspond all the meridian quadrants joining the poles to the equatorial points $\varepsilon' \cdot \varepsilon^\nu$, which thus bisect the angles between the meridian quadrants which we obtained before.* In particular, for $n = 6$, they make up the three entire meridians into which the real meridian is transformed by rotations of 30°, 90°, 150°.

There remain to be found the $2n$ curvilinear segments which correspond to the long-dotted half-meridian $-1 < w < +1$. I shall prove that they are *the segments of the equator of the z sphere determined by the points ε^ν and $\varepsilon' \cdot \varepsilon^\nu$.* In fact, the equator represents the points of absolute value one and is given therefore by $z = e^{i\varphi}$ where φ is real and ranges from 0 to 2π. Hence we have

$$w = \frac{1}{2}\left(z^n + \frac{1}{z^n}\right) = \frac{1}{2}\left(e^{ni\varphi} + e^{-ni\varphi}\right) = \cos n\varphi.$$

This expression is always real, and its absolute value is not greater than 1. In fact, it assumes once every value between $+1$ and -1 as φ varies from one multiple of π/n to the next one, i. e., when z traverses one of the segments of which we are speaking.

The curves determined in this manner divide the z sphere into $2 \cdot 2n$ triangular half-regions which are bounded by one curve of each of the three sorts, and each half-region corresponds to a half leaf of the Riemann surface. Several regions can meet only at the critical points, and then in accordance with the table of multiplicities (p. 116), namely, *$2n$ at the north pole, and at the south pole, and $2 \cdot 2$ at each of the points ε^ν and $\varepsilon' \cdot \varepsilon^\nu$.* In order to determine which of these regions are to be shaded, we notice that when w traverses, in order, the full-drwan, the long-dotted, and the short dotted parts of the real w meridian, the rear half of the w sphere lies at its left. Since the mapping is conformal with preservation of angles, we should shade those half-regions whose boundaries follow one another in this same sense, and we should leave the others unshaded.

We have now obtained a complete geometric picture of the mutual dependence between z and w which is set up by our equation. We might follow it out in greater detail by examining more closely the conformal mapping of the single triangular regions upon the w hemisphere, but we shall forego this. *I shall describe only,* and *briefly, the case $n = 6$, to which I have already given special attention.* The z sphere is then divided into twelve shaded and twelve unshaded triangles of which six of each sort are visible in Fig. 44. Six of each sort meet at each pole, and two of each sort at each of twelve equidistant points of the equator. Each triangle is mapped conformally upon a w half-leaf of the same sort. Of the half-leaves of the Riemann surface, six of each sort are connected at the branch position ∞, and two of each sort at each of the branch positions ± 1, corresponding to the grouping of the half-regions on the z sphere.

We may obtain a convenient picture of the division of the z sphere, and one which is especially valuable because of its analogy with pictures soon to come, as follows. If we join the n equidistant points on the equator (e. g., the ε^ν) with one another in order by straight lines, and also join each of them to the two poles, one obtains a *double pyramid, with $2n$ faces, inscribed in the sphere* (in Fig. 44, twelve faces). If we now project, from the center, the subdivision of the z sphere upon this double pyramid, every pyramid face is divided into a shaded and an unshaded half by the altitude of that face dropped from the pole. If we represent the division of the z sphere, and consequently our function, by means of this double pyramid, the latter will render a service quite analogous to that which we shall get in the coming examples from the *regular polyhedra.* We obtain a *complete analogy if we think of the double pyramid as collapsed into its base,* and consider the *double regular n-gon*

(hexagon) which results whose two faces (upper and lower) are divided each into $2n$ triangles by the straight lines which join the center with the vertices and the middle points of the sides (see Fig. 45). *I have been in the habit of calling this figure a dihedron and of classing it with the five regular polyhedra which have been studied since Plato's time.* It fulfills, in fact, all the conditions by means of which a regular polyhedron is usually defined, since its faces (the two faces of the n-gon) are congruent regular polygons, and since it has congruent edges (the sides of the n-gon) and congruent vertices (the vertices of the n-gon). The only difference is that it does not bound a proper solid body but encloses the volume zero. Thus the theorem of Plato, that there are

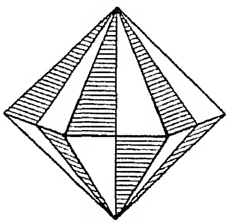


Fig. 44.

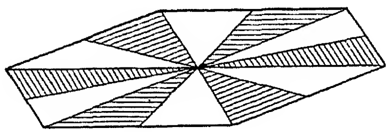


Fig. 45.

only five regular solids, is correct only when one includes in the definition the requirement of a *proper* solid, which is always tacitly assumed in the proof.

If we start with the dihedron, we obtain our subdivision of the z sphere by projecting upon that sphere not only its vertices but also the centers of its edges and its faces, the projecting rays for the latter being perpendicular to the plane of the dihedron. *Thus the dihedron can also be looked upon as representing the functional relation which our equation sets up between w and z.* Hence the brief name which we have already used, dihedral equation, is appropriate.

In addition, we shall now consider those equations which, as already intimated, are closely related to the platonic regular solids.

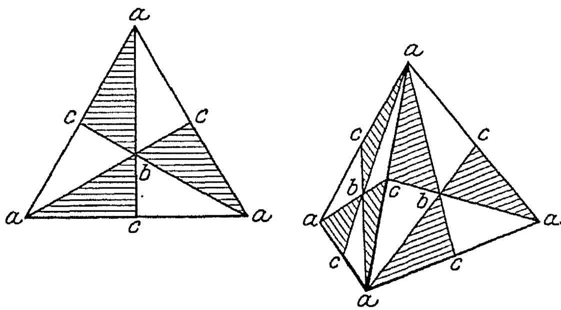
3. The tetrahedral, the octahedral, and the icosahedral equations.

We shall see that the last two could, with equal right, be called the hexahedral and the dodecahedral equations, so that all five regular bodies will have been covered. We shall follow here a route that is the reverse of the one we followed in the preceding example. *Starting from the regular body, we shall first deduce a division of the sphere into regions, and we shall then set up the appropriate algebraic equation, for which that figure* is the proper *geometric interpretation.* I shall have to confine myself frequently to suggestions, however, and I therefore refer you at once to my book: *Vorlesungen über das Ikosaeder und die Auf-*

lösung der Gleichungen vom fünften Grade[1], in which you will find a systematic presentation of the entire extensive theory with its numerous relations to allied fields.

Moreover, I shall give a parallel treatment of all three cases and I shall begin by *deducing the subdivision of the sphere for the tetrahedron.*

1. *The tetrahedron* (see Fig. 46). We divide each of the four equilateral face-triangles of the tetrahdron, by means of the three altitudes. into *six partial triangles.* These are congruent in two groups of three each, while any two non-congruent ones are symmetric. We obtain thus a *division of the entire surface of the tetrahedron into twenty-four triangles, which fall into two groups, each containing twelve congruent triangles, while any triangle of one group is symmetric to every triangle of the other group.* We shall shade the triangles of one group. Among the vertices of these twenty-four triangles we can distinguish *three sorts,* such that each triangle has one vertex of each sort:



Face Triangle (actual size). Tetrahedron.

Fig. 46.

a) *the four vertices of the initial tetrahedron, at each of which three shaded and three unshaded triangles meet*;

b) *the four centers of gravity of the faces,* which determine again another regular tetrahedron (the co-tetrahedron); *at each of these, three triangles of each kind meet*;

c) *the six middle points of the edges,* which determine a *regular octahedron*; *at each of these, two triangles of each kind meet.*

If from the center of gravity of the tetrahedron we project this subdivision into triangles upon the circumscribed sphere, the latter will be *subdivided into $2 \cdot 12$ triangles, which are bounded by arcs of great circles and are mutually congruent or symmetric.* About each vertex of the sort a), b), c), there will be respectively 6, 6, 4 equal angles, and since the sum of the angles about a point on a sphere is 2π, *each of the spherical triangles will have an angle $\pi/3$ at a vertex of the sort a or b and an angle $\pi/2$ at a vertex of the sort c.*

It is a *characteristic property of this division of the sphere that it, as well as the tetrahedron itself, is transformed into itself by a number*

[1] Leipzig 1884; referred to hereafter as "*Ikosaeder*". Translation into English by G. C. Morrice: *Lectures on the Icosahedron by Klein.* Revised Edition, 1911, Kegan Paul & Co.

of rotations of the sphere about its center. This will be clear to you in detail if you examine a model of the tetrahedron with its divisions, like the one in our collection. For the lecture, it will suffice if I indicate the *number of possible rotations* (whereby the position of rest is included as the *identical rotation.* If we select a definite vertex of the original tetrahedron, we can, by means of a rotation, transform it into every vertex of the tetrahedron (including itself), which gives four possibilities. If we keep this vertex fixed, however, in any one of these four positions, we can still transform the tetrahedron. This gives *altogether* $4 \cdot 3 = 12$ *rotations* which transform the tetrahedron, or the corresponding triangular division of the circumscribed sphere, into itself. By means of these rotations we can transform a preassigned shaded (or unshaded) triangle into every other shaded (or unshaded) triangle, and the particular rotation is determined when that second triangle is chosen. These *twelve rotations* form obviously what one calls *a group* G_{12} of twelve operations, i. e., if we performs two of them in succession, the result is one of the twelve rotations.

If we think of this sphere as the z sphere, each of these twelve rotations will be represented by a *linear transformations* of z, and *the twelve linear transformations which arise in this manner will transform into itself the equation which corresponds to the tetrahedron.* For purposes of comparison, I remark that *one can interpret the* $2n$ *linear substitutions of the dihedral equation as the totality of the rotations of the dihedron into itself.*

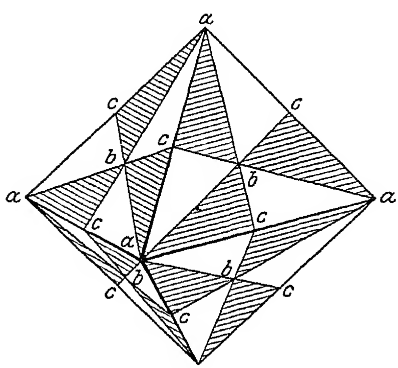


Fig. 47.

2. We shall now treat the *octahedron* similarly (see Fig. 47) and we may be somewhat briefer. We divide each of the faces, just as before, into *six partial triangles* and obtain a *division of the entire surface of the octahedron into twenty-four congruent shaded triangles, and twenty-four unshaded triangles which are congruent among themselves but symmetric to the other twenty-four.* We can again distinguish *three sorts of vertices*:

a) *the six vertices of the octahedron, at each of which four triangles of each kind meet;*

b) *the eight centers of gravity of the faces,* which form the *vertices of a cube*; at each of these, three triangles of each kind meet;

c) the twelve mid-points of the edges, at each of which two triangles of each kind meet.

If we pass now to the *circumscribed sphere*, by means of central projection, we obtain a division into $2 \cdot 24$ spherical triangles which are either congruent or symmetric, and each of which has an angle $\pi/4$ at the vertex a, $\pi/3$ at the vertex b, and $\pi/2$ at the vertex c. Since the vertices b form a cube, it is easy to see that *one would have obtained the same division on the sphere if one had started with a cube and had projected its vertices, and the centers of its faces and edges, upon the sphere.* In other words, we do not need to give special attention to the cube.

Just as in the previous case, it is easy to see *that the octahedron, as well as this division of the sphere, is transformed into itself by twenty-four rotations which form a group G_{24}; again each rotation is determined in that it transforms a preassigned shaded triangle into another definite shaded triangle.*

3. We come finally to the *icosahedron* (see Fig. 48). Here, also, we start with the same subdivision of each of the twenty-four triangular *faces and obtain altogether sixty shaded and sixty unshaded partial triangles.* The *three sorts of vertices* are:

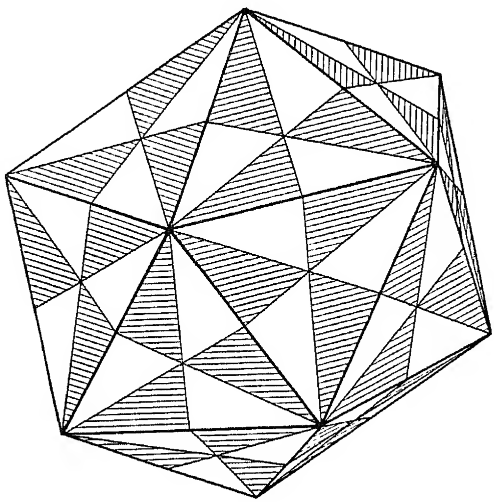


Fig. 48.

a) *the twelve vertices of the icosahedron, at each of which five triangles of each kind meet;*

b) *the twenty centers of gravity of the faces,* which are the *vertices of a regular dodecahedron*; *at each of them three triangles of each kind meet*;

c) *the thirty mid-points of the edges, at each of which two triangles of each sort meet.*

When this is carried over to the sphere each spherical triangle has at the vertices a, b, c the angles $\pi/5$, $\pi/3$, $\pi/2$, respectively. From the property of the vertices b one can conclude, as before, *that the same division of the sphere would have resulted if one had considered the dodecahedron.*

Finally, the icosahedron, as well as the corresponding division of the sphere, is transformed into itself by a group G_{60} of sixty rotations of the sphere about its center. These rotations, as well as those for the octahedron, will become clear to you upon examination of a model.

Let me make a list of *the angles of the spherical triangles* which have appeared in the three cases which we have considered, to which I shall add the dihedron also; they are

Dihedron:	$\pi/2$,	$\pi/2$,	π/n;
Tetrahedron:	$\pi/3$,	$\pi/3$,	$\pi/2$;
Octahedron:	$\pi/4$,	$\pi/3$,	$\pi/2$;
Icosahedron:	$\pi/5$,	$\pi/3$,	$\pi/2$.

As a variation of a joke of Kummer's I might suggest that the student of natural science would at once conclude from this, that there were additional subdivisions of the sphere, having analogous properties, and with angles such as $\pi/6$, $\pi/3$, $\pi/2$; $\pi/7$, $\pi/3$, $\pi/2$. The mathematician, to be sure, does not risk making such inferences by analogy, and his cautiousness justifies itself here, for *the series of possible spherical subdivisions of this sort ends*, in fact, *with our list*. Of course this is connected with the fact *that there are no more regular polyhedrons*. We can see the ultimate reason in *a property of whole numbers*, which does not admit a reduction to simpler reasons. It appears, namely, that the angles of each of our triangles must be aliquot parts of π, say π/m, π/n, π/r, such that the denominators satisfy the inequality

$$1/m + 1/n + 1/r > 1 .$$

This inequality has the property of existing only for the integral solutions given above. Moreover, we can understand it readily, since it only expresses the fact that the sum of the angles of a spherical triangle exceeds π.

I should like to mention that, as some of you doubtless know, an *appropriate generalization* of the theory does carry one byeond these apparently too narrow bounds: *The theory of automorphic functions* involves subdividing the sphere into *infinitely many triangles* whose angle sum is less than or equal to π.

4. Continuation: Setting up the Normal Equation.

We come now to the second part of our problem, *to set up that equation of the form*

$$(1) \qquad \varphi(z) - w\psi(z) = 0, \quad \text{or} \quad w = \frac{\varphi(z)}{\psi(z)},$$

which belongs to a definite one of our three spherical subdivisions, that is, which maps the two hemispheres of the w sphere upon the $2 \cdot 12$, or the $2 \cdot 24$, or the $2 \cdot 60$ partial triangles of the z sphere. To each value of w there must correspond then, in general, 12, 24, 60 values, respectively, of z, each one in a partial triangle of the right kind. Hence the desired equation must have the *degree 12, 24, 60* in the three cases respectively, for which we shall write N in general. Now each partial region touches